NUCLEIC ACID ASSEMBLY USING SMALL MOLECULE INTERACTIONS

A Thesis

Presented to

The Academic Faculty

by

Swapan Satyen Jain

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in Biochemistry

Georgia Institute of Technology

August 2006

NUCLEIC ACID ASSEMBLY USING SMALL MOLECULE INTERACTIONS

Approved By:

Dr. Nicholas V. Hud, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. W. David Wilson
Department of Chemistry
Georgia State University

Dr. Donald F. Doyle
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Loren D. Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. L. Andrew Lyon
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: July 08, 2006

Anaximander's views on the origin of man apply equally to the origin of chickens. Eggs need to be hatched and chicks need to be reared. Therefore, some nonchickens must have served as a parent. Consequently, there was a chicken egg before there were any adult chickens.

A Brief History of the Paradox: Philosophy and the Labyrinths of the Mind -: Sorensen

DEDICATION

I would like to dedicate this thesis to my dear bade papa (Sagar C. Jain). The way you have led your life to this day continues to be an inspiration to many of us. You have always been the beacon of love and hope for me. Thank you for always lighting the path amidst uncertainity and despair. I sincerely appreciate your profound love and sense of concern towards me. Thank you for teaching me the true meaning of the following quote:

"*What is done out of love always takes place beyond good and evil*" - Nietzsche

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Nicholas Hud, for the enormous role he has played in my education and overall growth in the past five years here at Georgia Tech. Thank you for teaching me to appreciate the importance of each experimental result and to let the data speak for itself. Even when the experiments do not go as planned, thank you for finding the 'good news' from each situation and in this manner, instilling a great deal of confidence and vigor in us to continue despite the uphill task ahead.

I am grateful to the past and present members of the Hud laboratory for creating an amazing environment for research. Time flew by in the company of Christine, Igor, Tumpa, Jim, Ozgul, Heather, Cathy, Eric, Matzaj, Christopher and many others who are not mentioned by name. You have been and always will be my dear friends and family away from home. I am also thankful to Sriram Kanvah and Nadia Boguslavsky for their assistance in our work.

I am unable to express my gratitude for my late grand parents (Mela Ram Jain and Ram Pyari Jain), bade mummy (Usha Jain), and my parents, Sumangla and Satya Paul Jain. Thank you for your everlasting love and support from thousands of miles away. I am also deeply indebted to my parents away from home, Tarsem and Beverly Singh. Without you, this would all still be a dream. Vicky bhaiya, thank you for your numerous trips to Atlanta and daily 9:15 pm phone calls. I appreciate the efforts of Dicky, Jennifer, and Salman for enduring with me during times when it was most difficult to do so. To Ritu didi, Manu bhaiya, Aarti didi and all other family and friends here in US and in India, I love you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

Acronym or Symbol	Definition
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
bp	base pair
A	adenine
T	thymine
G	guanine
C	cytosine
U	uracil
UV	ultraviolet
CD	circular dichroism
λ	wavelength (nm)
θ	ellipticity
$[\theta]$	molar ellipticity
$\Delta\varepsilon$	circular dichroism
ε_L	molar extinction coefficient of left circularly polarized light
ε_R	molar extinction coefficient of right circularly polarized light
A_L	absorbance of left handed polarized light
A_R	absorbance of right handed polarized light

mdeg	millidegrees (ellipticity units)
T_m	melting temperature
ΔG	change in Gibbs free energy
ΔH	change in enthalpy
ΔS	change in entropy
R	universal Gas Constant
I	incident light
I_0	transmitted light
rms	root mean square
PAGE	polyacrylamide gel eletrophoresis
ESI	electrospray ionization
MS	mass spectrometry
HPLC	high performance liquid chromatography
TLC	thin layer chromatography
M/Z	mass to charge ratio
E_a	activation energy
LC	ligation complex
K_d	equilibrium dissociation constant
TEAA	triethyl ammonium acetate
NNEP	nearest neighbor exclusion principle

SUMMARY

Research efforts in our lab are focused, in part, on the assembly and chemical coupling of nucleic acids by non enzymatic methods. In contemporary life, proteins and nucleic acids are intricately dependent upon each other for a host of functions including, but not limited to, replication and chemical ligation. Protein enzymes are necessary for the synthesis of DNA and RNA, while nucleic acids are necessary for both the coding and synthesis of proteins. According to the RNA World hypothesis, early life used nucleic acids for both information storage and chemical catalysis before the emergence of protein enzymes. However, it still remains a mystery how nucleic acids were able to assemble and replicate before the advent of protein enzymes. We have utilized the ability of small molecule intercalation to assemble nucleic acids into stable secondary structures. Our motivation in this pursuit comes from the recently proposed 'Molecular Midwife' hypothesis where small molecules may have acted as nanoscale structural scaffolds upon which the nucleic acid bases were able to stack into stable structures and undergo assembly into polymers.

Coralyne is a small, planar, crescent-shaped molecule which preferentially intercalates DNA triplex structures over duplexes. We have shown that nucleic acid binding by coralyne is a powerful means to control DNA secondary structure. Coralyne is able to disproportionate (repartition) $(dA)_n$•$(dT)_n$ duplex into $(dT)_n$• $(dA)_n$•$(dT)_n$ triplex and homo $(dA)_n$ self-structures. Formation of a stable homo $(dA)_n$ provides an excellent illustration of how small molecule intercalation can be used to drive nucleic acid assembly, because $(dA)_n$ does not form a stable multi-stranded self-structure at neutral

pH in the absence of coralyne. We have also found that the kinetics and thermodynamics of coralyne-drive duplex disproportionation and triplex stabilization are strongly dependent upon oligonucleotide length.

We have experimentally shown that small molecule binding to nucleic acids can facilitate template-directed synthesis of nucleic acids. Proflavine, a planar tricyclic cationic molecule is a classical duplex intercalator from the acridine family of dyes. Our work demonstrates that proflavine can assemble modified forms of the short oligonucleotides, $(dT)_3$ and $(dT)_4$, alongside a $(dA)_n$ template into a ligation active complex where the short oligomers become covalently linked together to form a $(dT)_7$ product. We observe greater than three orders of magnitude enhancement in product formation by simple addition of proflavine to the reaction mixture. An in-depth analysis shows that proflavine has multiple modes of binding to the ligation active complex and that some of these binding modes are deleterious to assembly. We also demonstrate that intercalation-mediated synthesis can promote ligation with non-Watson-Crick base pairs, which has implications regarding the creation of artificial self-replicating systems with non-natural base pairs.

Our results show that coralyne can assemble substrate strands into a ligation active complex based upon the triplex secondary structure. Finally, an efficient solution to the protein-free coupling of monomer bases into oligomeric structures has been sought for decades. Our work with proflavine has shown that assembly and coupling of monomers into larger order structures can be accomplished by small molecules at submicromolar concentrations of nucleic acids. These findings provide further credence to the robust and versatile nature of small molecule mediated assembly of nucleic acids.

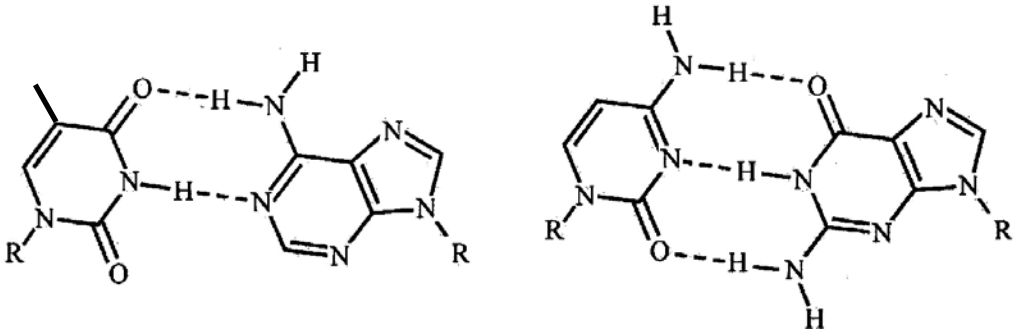
CHAPTER 1

INTRODUCTION

1.1. DISCOVERY OF DNA AS THE GENETIC MATERIAL

It is an indisputable fact of life that all living entities depend upon Deoxyribonucleic Acid (DNA) and Ribonucleic Acid (RNA) for their survival and function. Our present day understanding of nucleic acids comes from research efforts within the last 60 years. Transformation experiments by Avery using pneumococcal bacterium showed DNA to be the carrier of genetic information [1]. Hershey & Chase later demonstrated that bacteriophage DNA was found in the cell fraction of bacteria whereas the sulfur labeled proteins were found in the supernatant [2]. The discovery of pairing rules by Chargaff, where the percentage of purine and pyrimidine content in cells was 50% each, ultimately lead to our understanding that Adenine pairs with Thymine and Guanine pairs with Cytosine (Figure 1.1) [3]. These findings paved the way for Watson and Crick to propose a double helical model as the structure of DNA. In this model, the authors showed that DNA was composed of two antiparallel, complimentary strands that are held together by hydrogen bonding between the purine and pyrimidine base pairs [4].

Nucleic acid bases in each strand are attached to a sugar phosphate backbone which extends on the outside of the helix to minimize repulsions between the negatively charged phosphate groups [4]. B-DNA is a right handed, twisted (36º/bp) structure with 10 base pairs per turn of the double helix. Each turn of the helix is 34 Å in length which yields 3.4 Å rise per bp (distance between base pairs). Bases in DNA are attached to deoxyribose sugar (C2'-endo) with an anti orientation of the glycosidic bond [5].



A•T base pair G•C base pair

Figure 1.1. Chemical structures of Watson-Crick base pairs.

All living systems depend upon nucleic acid molecules for the storage and transfer of genetic information. Therefore, in order to understand life's origins, one must understand the origin of nucleic acids. Early experiments by Miller-Urey found that building blocks of proteins (amino acids) were synthesized by passing electric current through a model reducing prebiotic atmosphere of methane (CH_3), ammonia (NH_3), hydrogen (H_2) and water (H_2O) [6]. In 1961, Juan Oro showed that a significant amount of the nucleic acid base adenine could be produced under primitive Earth conditions from a mixture of hydrogen cyanide (HCN) and ammonia [7, 8]. These experiments had immense biological relevance because adenine is a building block of nucleic acids and ATP is the chief energy molecule in cells. Experiments have now shown that prebiotic synthesis of pyrimidines, cytosine and uracil, can also be accomplished. In an evaporating environment, cyanoacetaldehyde reacts with concentrated urea to give cytosine in 30 - 50% yield which can then be hydrolyzed to uracil [9]. It has also been shown that a reducing environment is not necessary for prebiotic synthesis of biologically relevant molecules. Miller *et al.* have shown that amino acids, bases, and base analogues

can be synthesized in a CO-rich environment [10]. Even the synthesis of adenine in an ammonia-free environment could have taken place by the photochemical conversion of a hydrogen cyanide tetramer in a eutectic environment [11].

The above experiments demonstrated that prebiotic synthesis of relevant biological molecules could be carried out. But living systems in contemporary life depend upon a close working relationship between proteins and nucleic acids for proper function. Proteins are necessary for the synthesis of nucleic acids whereas nucleic acids are vital for coding and synthesis of proteins. This presents a chicken and egg dilemma as to which polymers came first.

1.2. RNA WORLD HYPOTHESIS

Protein enzymes are the macromolecular machines of cells and inherently very complex structures with tertiary and quaternary contacts. It is now generally believed that the origin of nucleic acids preceded the origin of proteins. The chicken and egg dilemma can be answered by the 'RNA World Hypothesis' which proposes that an early form of life used RNA polymers for information storage as well as catalytic function [12]. The remarkable ability of RNA oligomers to fold into tertiary structures and catalyze protein-free self-splicing in the presence of divalent cations provided one of the first experimental evidence for the RNA World hypothesis [13]. It has been shown that the ability of RNA to act as efficient templates simplifies non-enzymatic information transfer and self-replication [14, 15]. The feasibility of RNA catalyzed RNA polymerization has been illustrated by the splicing activity of multiple oligonucleotides on a *Tetrahymena* ribozyme template [16, 17]. The self-cleavage activity of the hammerhead ribozyme also

illustrates that functions carried out by enzymes in contemporary living systems could have been accomplished by RNA in early life [18]. Bartel's demonstration that a single RNA molecule can fold into several structures and catalyze multiple reactions provided further credence to the idea that RNA could have been the first molecule capable of carrying genetic information as well as catalyzing reactions [19]. The finding that RNA molecules can control gene expression by acting as riboswitches and ribozymes adds to the repertoire of RNA as a molecule of enormous capabilities [20-26]. In addition to RNA-mediated catalysis, a number of laboratories have now shown that DNA also possesses catalytic activity. *In vitro* selection of deoxyribozymes have led to the evolution of DNA sequences with self-splicing, ligase, kinase and phosphorylation activities [27-31]. In addition, synthesis of native 3'-5' RNA linkages and RNA ligation can be carried out by DNAzymes [32, 33]. All of the above mentioned experiments rely on the presence of RNA polymers. A simple question can be asked: Where did the RNA polymers come from?

1.3. TEMPLATE-DIRECTED CONDENSATION REACTIONS

Despite the ability of RNA polymers to carry out self catalysis, it still remains a mystery how RNA, or an ancestor of RNA, was able to assemble and replicate before the advent of protein enzymes. Significant gaps in our understanding remain regarding the emergence of RNA polymers. The RNA World hypothesis can only be plausible in the presence of RNA polymers. Therefore, making a transition from nucleic acid monomers to polymeric structures is necessary. However, this process remains poorly understood.

Exhaustive efforts to condense individual nucleotides alongside nucleic acid templates have been only marginally successful [34]. For example, experiments relying solely on chemical activation to improve the rate of spontaneous mononucleotide coupling along existing templates have proven rather inefficient. Orgel's group has shown that activated guanosine monomers can be oligomerized on Poly(C) template strands. However, these experiments have been restrictive in the choice of substrates and in addition, they suffer from poor efficiency. Poly(G) templates cannot be used to condense activated cytosine bases perhaps because of the formation of G-quadruplex structures and poor stacking of cytosines. Similarly, thymine bases stack poorly and therefore cannot be utilized efficiently in condensation along a Poly(A) template [35-39]. Cyanogen bromide chemistry has also been used to achieve template-directed polymerization of oligoadenylates but the reactions suffer from low yields [40]. In the presence of montmorillonite clay, template-directed as well as template independent oligomerization of activated monomers has been shown by the work of Ferris and coworkers but a good understanding of the mechanism by which minerals present in the clay catalyze these reactions is lacking [41-45]. Chain growth polymerization by reading a template sequence has been accomplished by using reductive amination chemistry but product formation is limited to short oligonucleotides [46-50]. Based upon such research, it is now generally accepted that base pair interactions alone are not sufficient for the spontaneous assembly of nucleic acid monomers into higher order polymers [51]. The activation chemistries used are non natural and more importantly, these experiments do not answer the question of the emergence of templates.

1.4. SELF REPLICATION AND AUTOCATALYSIS

The ability to design a molecular system that utilizes template-directed synthesis for chemical information transfer without the aid of enzymes represents a major goal for chemists trying to develop a system capable of replication. A self-replicating system is able to catalyze the formation of additional copies of itself by utilizing components in solution. Autocatalytic behavior gives rise to and is evidenced by the exponential product growth. Despite several decades of research, only limited success has been achieved in this endeavor [52]. The first successful example of a non-enzymatic, self replicating system comes from the work of von Kiedrowski and coworkers where it was shown that autocatalytic synthesis of a self-complimentary hexanucleotide can be accomplished using two complementary trinucleotides [53].

Rebek's laboratory has also demonstrated autocatalysis in a minimalistic system where an amino adenosine triacid ester (AATE) molecule makes additional copies of itself by templating the coupling of an ester and an amino adenosine molecule [54]. Nicolau's laboratory has shown that self-replication of a palindromic duplex can be achieved via the formation of a triplex intermediate. This process is sequence selective, as its efficiency suffers in the presence of base pair mismatches [55]. Finally, Ghadiri and coworkers have demonstrated that a 32-residue alpha helical peptide can template additional copies of itself by coupling a 15-residue and 17-residue fragment. The requirement that the self replication process display parabolic growth is shown in their work as the initial rate of product formation corresponds to the square-root of initial template concentration [56].

1.5. MOLECULAR MIDWIFE HYPOTHESIS

Previous work in the literature has shown that RNA polymers can possess catalytic activity and may also be capable of self replication. But a viable bridge from 'The Small Molecule World' to the 'The RNA World' is missing. A multitude of small molecules must have existed on prebiotic Earth, and only a small subset of these molecules is likely represented by the building blocks of biopolymers in contemporary life (e.g. sugars, amino acids, nucleotide bases). In 2000, Hud and Anet proposed a hypothesis which presents a new approach to the origin of nucleic acid polymers. The *molecular midwife* hypothesis proposes that small molecules that intercalate the bases of nucleic acids could have also been present on the prebiotic Earth and acted as nanoscale surfaces upon which nucleotide bases (RNA or RNA-like) stacked to form ordered assemblies (Figure 1.2) [57].

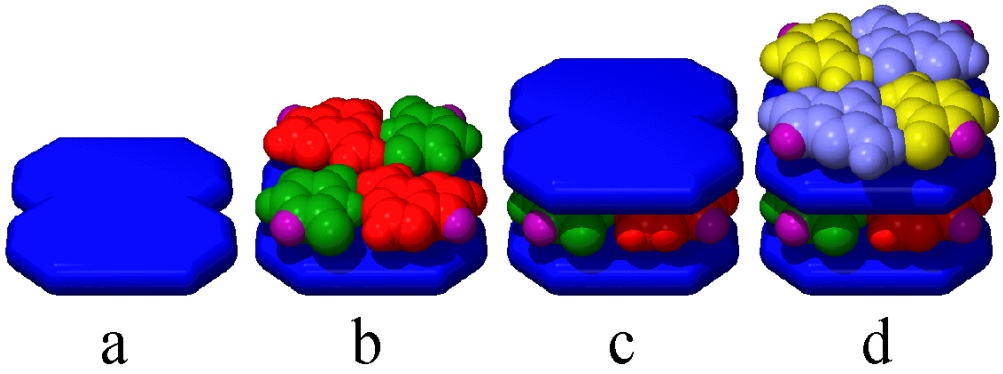


Figure 1.2. Schematic representation of the molecular midwife hypothesis. (a) A minimalistic representation of the midwife molecule. (b) Flat, planar surface of the midwife molecule allows for stacking of an $(A•U)_2$ tetrad (space-filling model). (c) Two midwife molecules are intercalated by an $(A•U)_2$ tetrad. (d) A $(G•C)_2$ tetrad is stacked on top of a midwife molecule which is stacked on top of an $(A•U)_2$ tetrad leading to the formation of columnar stacks. (Taken from reference [57]).

Furthermore, the formation of columnar stacks of midwife molecules that alternated with hydrogen-bonded bases could have provided a means for the formation and replication of the first RNA-like polymers. These assemblies are non-covalent and reversible structures that could have been coupled together by low energy bonds instead of the high energy phosphate linkages in present day nucleic acids [58]. Thus, the molecular midwife hypothesis represents a potential bridge between a prebiotic small-molecule world and the RNA World.

1.6. SMALL MOLECULE NUCLEIC ACID BINDING

The molecular midwife hypothesis presents a novel solution for the origin of nucleic acid polymers where small molecule intercalators could have assembled nucleic acid building blocks into higher order structures. In order to fully appreciate assembly of nucleic acids by small molecules, we must have a good understanding of the mechanisms by which small molecules interact with nucleic acids.

Structural understanding of a molecule is necessary in order to fully understand how it serves in its function. The rich structure of DNA lends itself to sequence specific recognition and regulation by a diverse array of molecules. Functional groups present in the major and minor groove of nucleic acids are a target of proteins and small molecules alike. Proteins interact with nucleic acids primarily by sequence specific recognition of the major groove. Electrostatic, hydrogen bonding and hydrophobic interactions are chiefly responsible for protein binding to nucleic acids.

Small molecules that interact with nucleic acid structures have been used as anticancer agents for decades [59-63]. A recent example comes from the work of Hurley laboratory, who have shown that porphyrin-like molecules can be used to stabilize G-quadruplex conformation in the *c-MYC* gene and control transcription regulation [64]. Antitumor activity of many small molecules comes from their ability to stabilize nucleic acid structures by sequence specific interactions. Normal functions, some of which include replication and transcription activity, are disrupted in the presence of stable drug-DNA complexes. Many of these compounds occur in nature and interact with DNA and RNA in a strong and sequence specific manner. Interactions of small molecules with nucleic acids are intricately tied to the structure of the small molecule.

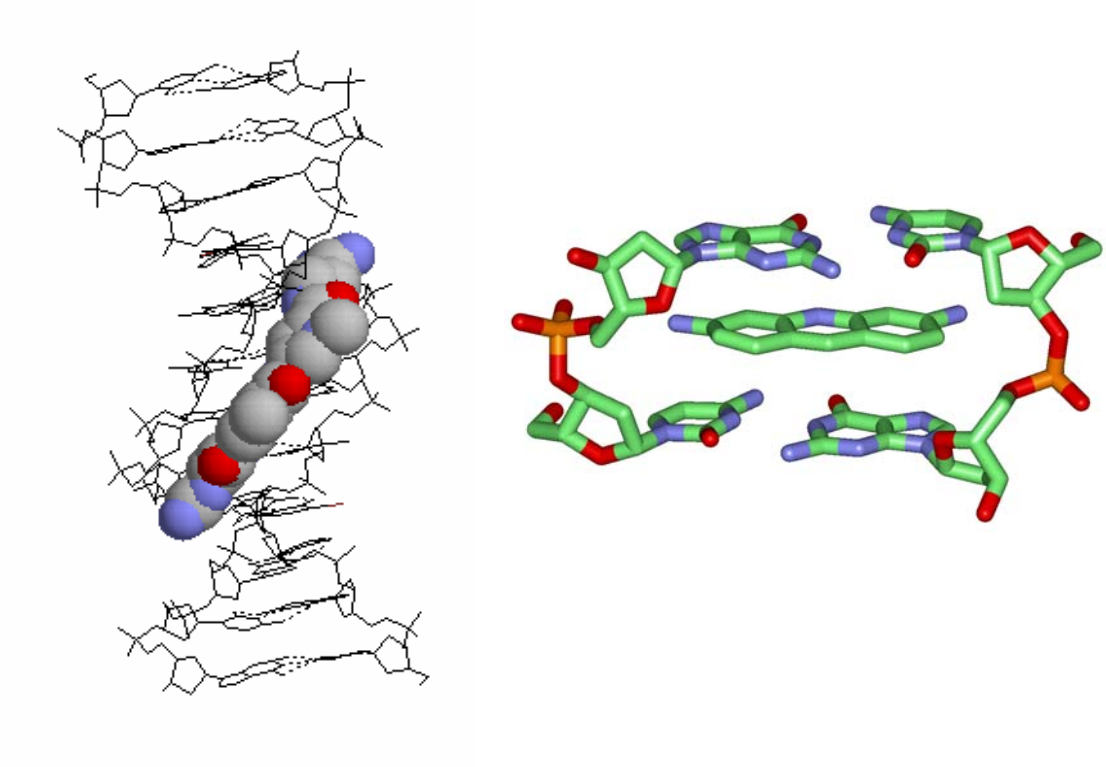


Figure 1.3. Minor groove binding and intercalation of nucleic acids. Minor groove bound Netropsin is shown on left [65]. Proflavine intercalating a G•C base pair is shown on the right [66].

1.6.1. Small Molecule Intercalation of Nucleic Acids

Intercalation of nucleic acids by small molecules was first proposed by Lerman in 1961 [67]. Significant changes in the solution viscosity and sedimentation coefficients of DNA were observed when acridine dyes such as proflavine and acridine orange were added to the sample solution. Based upon these structural changes in DNA, Lerman hypothesized that the ligands must be interacting with nucleic acids by a binding mode in which lengthening of the nucleic acid backbone takes place. Structural studies later confirmed that in the intercalation model, a geometrically flat, planar, aromatic and heterocyclic molecule is inserted between two base pairs of a double helix (Figure 1.3) [68]. Free energy of stacking between the small molecule and base pairs is the principal driving force behind intercalation. Small molecule intercalation leads to significant change in the structural parameters of nucleic acids. Two major changes that occur upon intercalation are lengthening and unwinding of the helix. The degree of helix lengthening and the magnitude of the unwinding angle of the double helix is chiefly dependent upon the structure and shape of the intercalating molecules.

One of the first examples of spectroscopic determination of the intercalation model came from the work of LePecq and Paoletti [69]. The fluorescence intensity of an ethidium-bound complex was found to be approximately 20-fold more intense than free ethidium in solution. A weaker binding process characterized by significantly reduced fluorescence intensity was also observed. This weaker binding process was found to be ethidium bromide interactions alongside the sugar phosphate backbone of DNA [69]. Besides insertion of planar, heterocyclic ring system of ligands between nucleic acid base pairs, there are other groups present on the ligand molecules that also interact with

phosphate and sugar molecules in the nucleic acid backbone. For example, ethidium bromide, daunomycin, actinomycin are molecules that intercalate DNA duplexes and contain exocyclic groups that interact in the grooves of DNA by hydrophobic, electrostatic, and hydrogen bonding interactions [62, 70-73].

1.6.2. Minor Groove Binding of Nucleic Acids by Small Molecules

Small molecules have also been recently developed that bind sequence specifically in the minor groove of duplex DNA (Figure 1.3) and function as artificial gene regulation elements [68]. The curvature of these molecules strongly resembles the helical curvature of the DNA minor groove (Figure 1.2). Ring polyamides from Dervan's group are a class of permeable molecules with proven great potential in artificial gene regulation [74, 75]. Hoechst 33258 is one of the most extensively studies minor groove binding agents where the bis-benzimidazole moiety is crucial in binding to the groove of nucleic acids [76]. Minor groove binding drugs such as Netropsin and Distamycin bind to A•T rich regions of DNA minor groove [65]. Distamycin has been shown to bind to the minor groove of DNA in a 1:1 as well as a 2:1 (drug:DNA site) binding ratio [77]. Minor groove agents, for the most part, specifically bind AA/TT sites spanning 4-5 base pairs. Molecules that bind in the DNA minor groove do so without minimal alteration of the helical structure of nucleic acids, which is opposite to the major structural perturbations caused by intercalation. Some of groove binding molecules are known to bind nucleic acid structures with nanomolar dissociation constants [65, 73, 77, 78].

Groove binding and intercalation are examples of noncovalent interactions of small molecules with nucleic acids. A number of molecules, synthetic and natural, also

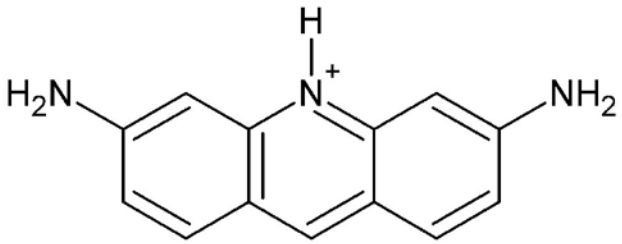
interact with nucleic acids in a covalent fashion. Bleomycins are a class of anticancer compounds that cause radical-mediated cleavage of DNA backbone in the minor groove [79-81]. In contrast, Cisplatin is an antitumor drug which leads to inter- and intra-strand adduct formation in the major groove of DNA [82]. In addition to duplex DNA, a number of laboratories have also worked towards the development of small molecules that specifically bind and stabilize triplex and G-quadruplex DNA [64, 83, 84]. The use of such molecules *in vivo* to facilitate the formation of non-duplex DNA structures has principally been pursued as a possible route to anti-gene therapy. As mentioned above, the recent demonstration that a cationic porphyrin derivative can suppress gene expression by promoting the formation of a G-quadruplex in living cells clearly illustrates the potential for this approach in the development of new therapeutics [64].

1.7. NUCLEIC ACID ASSEMBLY BY SMALL MOLECULE BINDING

The molecular midwife hypothesis proposes that small molecule binding to nucleic acids can be used to assemble nucleic acid structures. Experimental evidence for the molecular midwife hypothesis comes from recent work in our lab where we show that intercalation of nucleic acids by small molecules can be a very useful technique for the purpose of assembly and replication of nucleic acids [85-88]. We have focused on two small molecules for this study: proflavine (duplex assembly) and coralyne (triplex assembly).

1.7.1. Interaction of Proflavine with DNA Duplex Structures

Proflavine is a small planar tricyclic molecule amongst a group of molecules known to be classical duplex intercalating ligands (Figure 1.4) [63, 67, 89, 90]. Proflavine is a cationic molecule with a +1 charge at neutral pH (Figure 1.4). The pKa of the ionizable imino group is 9.1 [91]. Proflavine binding to nucleic acids is driven primarily by stacking interactions with the adjacent base pairs. Electrostatic interactions also drive proflavine binding to nucleic acids. Proflavine binds to duplex structures with a K_d of approximately 3 μM [89, 91]. Binding constants of small molecules to DNA can be affected by several different parameters. Bloomfield *et al.* demonstrated that proflavine binding to phage T2 DNA decreases both as a function of increasing ionic strength and temperature [92, 93]. Aggregation and dimerization behavior of geometrically flat and planar must be taken into consideration during experimental design. Proflavine has a reported dimerization constant of ca. 2 mM [94, 95].



Proflavine

Figure 1.4. Chemical structure of proflavine.

Viscosity studies have been routinely used to measure the changes in structural parameters of nucleic acids upon intercalation by small molecules. Intercalation of

nucleic acids by proflavine leads to an increase in contour length of DNA by 0.35 nm per drug molecule. This distance represents an increase in the DNA length by one bp spacing. In addition, proflavine intercalation leads to an unwinding of the helix by ~ 17 º [5].

In addition to intercalation, proflavine can also bind to nucleic acid structures by outside stacking. It has been reported that a weaker binding process involves the stacking of positively charged proflavine molecules alongside the negatively charged sugar-phosphate backbone of nucleic acid structures [90]. The weaker outside binding of molecules to nucleic acid backbones is not limited to proflavine as ethidium bromide has also been shown to interact with nucleic acids with dual binding modes [70].

Acridine dyes, such as proflavine, were of interest for the studies presented here because molecules have no bulky groups or moieties extending from the planar ring system of the molecules. We have exploited this property of proflavine, in addition to its flat structural match to Watson-Crick base pairs, for assembly of a duplex from two substrates and a template strand. In our work, we show that proflavine can assemble a stable ligation active complex composed of $(dT)_3$ and $(dT)_4$ substrates alongside a $(dA)_7$ template strand. This is a remarkable discovery considering that the equilibrium amount of substrate strands aligned along the template in solution is negligible. We have also uncovered the fact that proflavine assembly of a duplex structure is much more complicated than we previously appreciated. We find evidence for additional modes of proflavine binding to nucleic acids. These additional binding modes of proflavine, besides intercalation, can also facilitate assembly of duplex structures.

1.7.2. Assembly and Stabilization of Triplex Structures by Small Molecules

Formation of triplex structures was observed by Felsenfeld and Rich almost 50 years ago [96]. In the presence of divalent cations, it was found that a single strand of Poly(rA) binds to two Poly(rU) strands. It was later shown that heating in the presence of 200 mM NaCl converts an RNA duplex into a triplex structure [97]. Binding of a homopyrimidine strand by Hoogsteen pairing [98] into the major groove of a homopurine•homopyrimidine duplex leads to the formation of a DNA triplex. But the biological relevance of nucleic acid triplex structures has only been appreciated in the last two decades starting with the discovery of H-DNA structures [99, 100]. Many laboratories have focused their attention on stabilization of triple helical structures *in vivo* at physiological conditions as a potential tool in anticancer gene therapy [101-105]. An elegant example of triplex mediated duplex inhibition has been shown by Dervan and coworkers. In this work, the authors show that sequence specific DNA cleavage can be mediated by attaching Fe-EDTA to a triplex forming oligonucleotide (TFO) [106].

It has also been shown that polypurine/polypyrimidine hairpins can adopt triple helical structures at slightly acidic pH (i.e. pH 5) [107]. However, the stability of a Hoogsteen paired strand is appreciably lower than that of a Watson-Crick duplex [108]. Due to the lower stability of Hoogsteen hydrogen bonds, triplex based anticancer gene therapy has not gained traction as a formidable gene regulation technique. Small molecules that can stabilize triple helical structures may serve as potential anticancer agents. Stability of triplex structures can be increased by ligand induced stabilization similar to duplex structures. Many small molecules have been shown to stabilize triplex structures by binding in an intercalative mode [72, 103, 109-112].

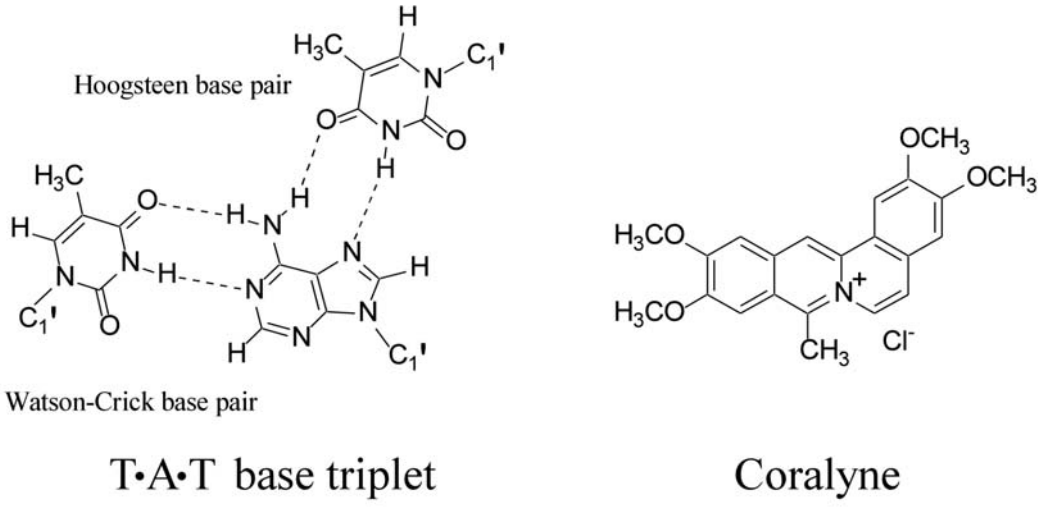


Figure 1.5. Chemical structures of a T•A•T base triplet (left) and coralyne chloride (right). The shape and structure of coralyne mimics closely to that of a base triplet.

Even ethidium bromide, a classical duplex intercalating species, binds to DNA triplexes by intercalation [102, 104]. On the other hand, groove binding molecules such as DAPI and Hoechst bind to the grooves of triplex structures [113].

Coralyne is a small, heterocyclic molecule that has a shape and structure similarity to that of a T•A•T base triplet (Figure 1.5). It is amongst a class of triplex-specific molecules and has been implicated in having potential antileukemic activity [112, 114]. Coralyne has been known to preferentially intercalate DNA triplexes over duplexes, and to increase the thermal stability of triplex DNA [109, 110, 115]. Data from a competition dialysis experiment by Chaires and coworkers presented in Figure 1.6 shows the relative binding of coralyne to several nucleic acid structures [72]. Triplex DNA shows the highest binding affinity amongst the group of nucleic acid structures.

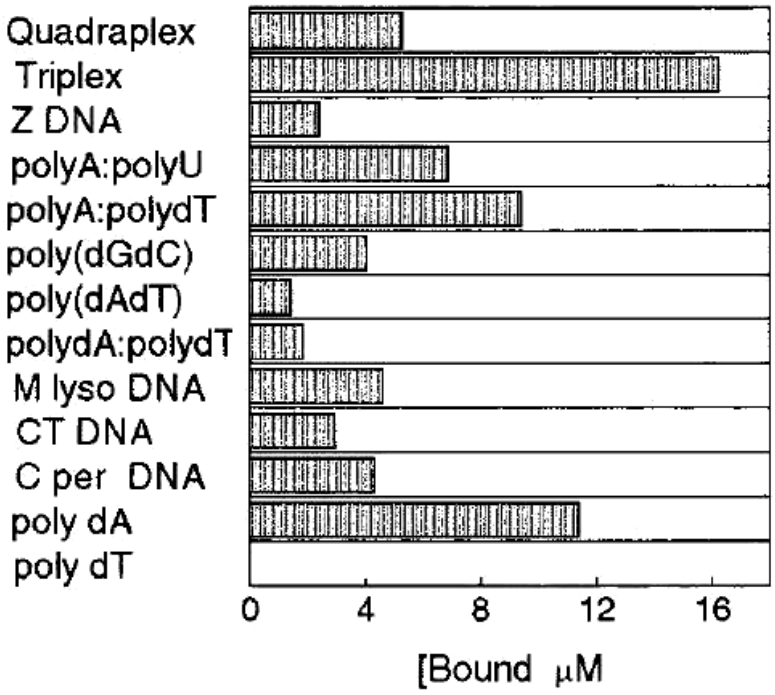


Figure 1.6. Competition dialysis study of coralyne binding to various nucleic acid structures. Figure taken directly from reference [72].

We have taken clues from the above studies and shown that coralyne can be used for the assembly and stabilization of triple helical structures. Specifically, we have shown that coralyne causes disproportionation of $(dA)_n \bullet (dT)_n$ duplexes into triplex $(dT)_n \cdot (dA)_n \bullet (dT)_n$ and single stranded $(dA)_n$ structures [86, 88]. We have also found that there are complex kinetic and thermodynamic factors that govern coralyne-mediated assembly and stabilization of nucleic acid structures.

1.7.3. Assembly and Stabilization of Novel Structures by Small Molecule Binding

Small molecules have been extensively used in the Hud laboratory for the non-covalent assembly of nucleic acids. Promotion of novel secondary structures by small molecules represents a powerful approach for self assembly and replication. Felsenfeld

and Stevens demonstrated that heating an RNA duplex led to the formation of a triple helix and single-stranded poly(rA) [97]. Research efforts from our laboratory have shown that a homo(dA) strand forms a stable secondary structure in the presence of coralyne. We have also shown that the homo(dA) self structure is an antiparallel duplex with A•A pairs [87]. Competition dialysis data shown in Figure 1.6 illustrates the remarkable ability of coralyne to bind to poly(dA) structures. It was also recently shown that coralyne forms a self structure in the presence of Poly (rA) single strands with a thermal stability of 60ºC [116].

Our work with proflavine has also shown that assembly of nucleic acids can be extended beyond Watson-Crick base pairing to artificial pairing systems that use non-standard base pairs. Proflavine can assemble secondary structures composed entirely of homo-thymine oligonucleotides. Thymine templated assembly and ligation of homo(dT) substrates is a novel finding and illustrates the enormous potential for the development of small molecule promoted formation of artificial self replicating systems.

1.8. SPECTROSCOPIC APPLICATIONS FOR UNDERSTANDING DRUG-DNA INTERACTIONS

Small molecule interactions with nucleic acids have been investigated by a variety of different techniques. While X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy provide atomic level resolution, spectroscopy can be used to study properties such as π-π stacking of bases. Ligands that interact with DNA base pairs can lead to marked changed in the electronic properties of nucleic acid bases; therefore techniques such as ultraviolet (UV) absorption and circular dichroism spectroscopy can

be particularly useful. Ligand binding to nucleic acids can also lead to shifts in the absorption maxima at specific wavelength which has also been used to assign the mode of small molecule binding to nucleic acids. Melting analysis by UV spectroscopy has been vital in understanding small molecule interactions with nucleic acids because many small molecules are known to greatly increase the thermal stability of DNA and RNA structures.

Circular dichroism (CD) spectroscopy provides information about the helical conformation of nucleic acids and therefore can be more useful than UV spectroscopy in certain circumstances. CD spectroscopy is based upon the absorption of circularly polarized light. Plane polarized light has two components: a left handed circular component and a right handed circular component. Upon passage of light through a chiral medium, left handed circularly polarized light absorbs at a different intensity than right handed circularly polarized light which leads to a differential absorption ($\varepsilon_L \neq \varepsilon_R$) between the two components of plane polarized light. Elliptical behavior observed in a spectrum is called 'The Cotton Effect'. Cotton effect can yield bands that can be positive (peak) or negative (trough) in intensity because circular dichroism is the change between two components of circularly polarized light. UV spectroscopy on the other hand only gives positive bands which are characterized as maxima and minima [117].

This differential absorption is referred to as circular dichroism ($\Delta\varepsilon$).

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R \tag{1.1}$$

Circular dichroism follows from Beer-Lambert's law. Beer's law dictates that the intensity of light passing though a material decreases exponentially,

$$I = I_0 10^{-\varepsilon cl} \tag{1.2}$$

where I is the intensity of transmitted light, I_0 is the intensity of incident light, c is molar concentration, l is path length in cm and ε is the molar extinction coefficient ($M^{-1}cm^{-1}$). Given that Transmittance (T) = I / I_0 and Absorbance (A) = log (I_0 / I), we have the following relationship:

$$A = \varepsilon c l \tag{1.3}$$

Circular dichroism is the absorption difference between the left circularly polarized and right circularly polarized incident light. Molar CD is defined by the following relationship:

$$\varepsilon_L - \varepsilon_R = \frac{(A_L - A_R)}{lc} \tag{1.4}$$

Ellipticity is an effect of circular dichroism where plane polarized light, upon absorption by a chiral medium, becomes elliptically polarized (Figure 1.7). Molar ellipticity [θ], in units of deg M^{-1} cm^{-1} * 100, is related to the CD ($\varepsilon_L - \varepsilon_R$) by the following relationship:

$$\theta = 3298 \times (\varepsilon_L - \varepsilon_R) \tag{1.5}$$

Ellipticity can be converted to molar ellipticity

$$[\theta] = \frac{\theta}{10 \times c \times l} \tag{1.6}$$

where [θ] is molar ellipticity, θ is ellipticity, c = concentration (M), and l = path length (cm).

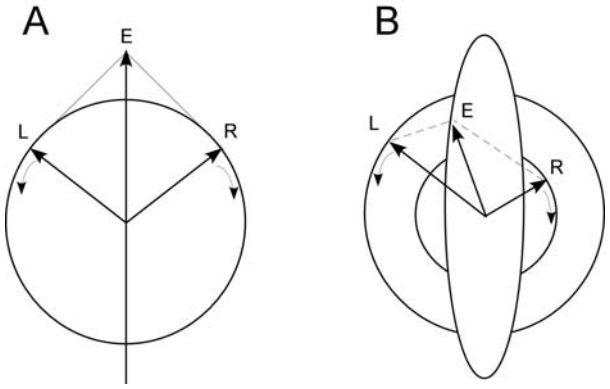


Figure 1.7. Illustration of circular dichroism and origin of ellipticity. (A) Plane polarized light is the sum of two components. L is left circularly polarized light and R is right circularly polarized light. If the intensity and phases of L and R are same, the resultant vector (E) lies in plane. (B) Upon passing through an optically active medium, the two components of plane polarized light can emerge with different intensities and phases. The resultant vector, E, follows an elliptical path which leads to elliptically polarized light. The angle, ψ, determined from the elliptical path is referred to as ellipticity with units in degrees. Adapted from Ref. [117].

CD spectroscopy can be a very powerful technique for investigation of small molecule interactions with nucleic acids. A large number of small molecules are achiral and therefore do not give CD signal because the two components of polarized light (left and right) are absorbed equally. However, upon interactions with nucleic acids, an induced CD effect is observed. Induced CD is defined as ellipticity observed in the absorption region of achiral guest chromophores. Stacking interactions (electronic transitions) between achiral ligands (guest) and chiral nucleic acids (host) can lead to induced CD [117].

Changes in melting temperature, modes of interactions, assembly of unstable structures, binding constants determination, and stoichiometry of ligand-nucleic acid binding are some of the drug-DNA parameters that we have been able to determine by using spectroscopic techniques.

CHAPTER 2

CORALYNE MEDIATED DUPLEX DISPROPORTIONATION AND TRIPLEX STABILIZATION

2.1. INTRODUCTION

For many years, small molecules that bind duplex DNA were primarily studied with the goal of understanding their activity in either the cause or treatment of cancer [59-63]. Within the last decade, the proposed applications for small molecule-DNA interactions have increased substantially. For example, small molecules have now been developed that bind sequence specifically in the minor groove of duplex DNA and function as artificial gene regulation elements [74, 75]. In addition to duplex DNA, a number of laboratories have also worked towards the development of small molecules that specifically bind and stabilize triplex and G-quadruplex DNA [64, 84, 118]. The use of such molecules *in vivo* to facilitate the formation of non-duplex DNA structures has principally been pursued as a possible route to anti-gene therapy.

Our laboratory is also interested in the use of small molecules as a general means to drive nucleic acid assembly and structural transitions. Part of our motivation in this pursuit comes from our recent proposal that small molecule intercalation may have played a central role in the early days of life (i.e. the RNA World) by facilitating nucleic acid assembly and replication [57]. In the same way, small molecule intercalation may provide a means to protein-free template-directed synthesis of nucleic acids [57]. Towards this end, we have initiated studies of how small molecule binding to nucleic acids by intercalation can drive the assembly of multistranded DNA and RNA structures.

Coralyne (Figure 2.1A) is small crescent-shaped molecule that is among a group of molecules known to preferentially intercalate DNA triplexes over duplexes, and to increase the thermal stability of triplex DNA [72, 109, 110, 112, 114].

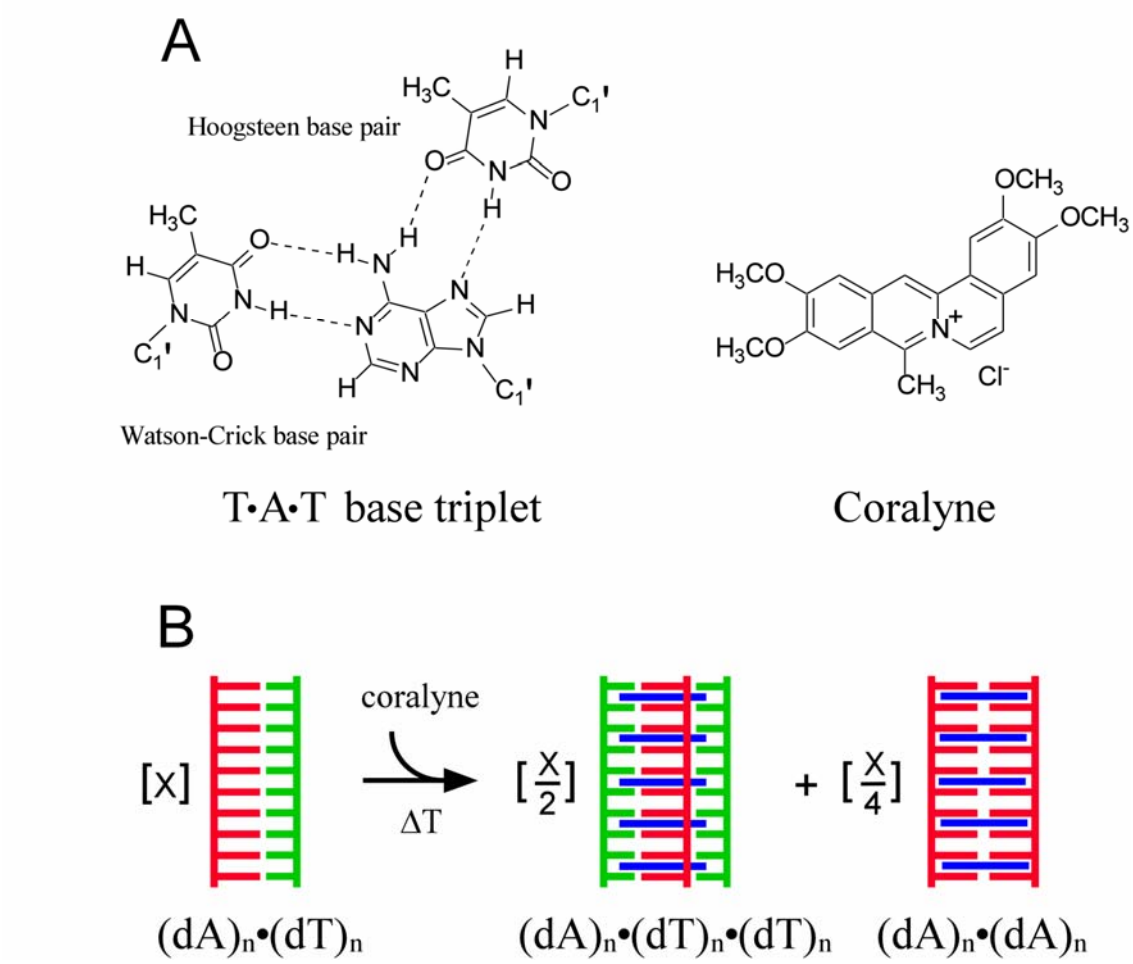


Figure 2.1. Coralyne-mediated assembly of a triplex and applicable molecular structures. (**A**) Structural representations of the T·A·T base triplet and coralyne chloride. (**B**) Schematic representation of the disproportionation of an $(dA)_n \cdot (dT)_n$ duplex into coralyne-intercalated triplex $(dA)_n \cdot (dT)_n \cdot (dT)_n$ and the $(dA)_n$ self-structure.

Previously, the Hud laboratory demonstrated that coralyne can drive the complete and irreversible disproportionation of duplex poly(dA)·poly(dT) [88]. That is, coralyne causes the repartitioning of duplex poly(dA)·poly(dT) into coralyne-intercalated triplex poly(dA)·poly(dT)·poly(dT) and poly(dA) (Figure 2.1B). We also discovered that poly(dA) forms a self-structure in the presence of coralyne that is stable up to at least 47°C. Data from several experimental techniques indicate that this poly(dA) self-structure is an antiparallel duplex with A·A base pairs that is intercalated up to a level of one coralyne molecule per two base pairs [87, 88]. This serendipitous discovery provides an excellent illustration of how small molecule intercalation can be used to drive nucleic acid assembly, because poly(dA) does not form a stable multi-stranded self-structure at neutral pH in the absence of coralyne.

Here, we show the effects of DNA strand length on the ability for coralyne binding to control DNA secondary structure [86]. Coralyne is shown to cause the complete disproportionation of duplex $(dA)_{16}\cdot(dT)_{16}$ at 36°C, as we previously demonstrated for duplex poly(dA)·poly(dT) [88]. However, our studies of duplex $(dA)_{16}\cdot(dT)_{16}$ have revealed that duplex disproportionation by coralyne is not strictly irreversible. Over the course of hours at 4°C, a disproportioned $(dA)_{16}\cdot(dT)_{16}$ sample reverts back to the duplex state from the coralyne-intercalated triplex and $(dA)_{16}$ self-structure. Furthermore, at room temperature a disproportioned duplex $(dA)_{16}\cdot(dT)_{16}$ sample requires days to reach equilibrium, and the equilibrium state contains a mixture of duplex, triplex and $(dA)_{16}$ self-structure. Coralyne-disproportioned samples of duplex $(dA)_{32}\cdot(dT)_{32}$ and duplex poly(dA)·poly(dT) were found to require several weeks to reach

structural equilibrium. We also show that even the kinetics of triplex stabilization by coralyne depend on oligonucleotide length [86].

We have also investigated the effects of C•G bp on coralyne-driven duplex disproportionation and triplex formation. Studies of coralyne binding with RNA duplex Poly r(A)•Poly(rU) and triplex Poly(rU) •Poly(rA) •Poly(rU) structures have also been carried out. We have also found that Poly(rA) binds coralyne similarly to Poly(dA). The results reported here demonstrate that the process of secondary structure stabilization by small molecule intercalation is potentially more complex than previously appreciated. Our results also illustrate the potential for small molecule binding to be used in concert with temperature to drive nucleic acid structural transitions.

2.2. EXPERIMENTAL PROCEDURES

2.2.1. Materials

Oligonucleotides $(dA)_{16}$, $(dT)_{16}$, $(dA)_{32}$, $(dT)_{32}$, $d(AAAGAAAG)_4$, $d(TTTCTTTC)_4$, $d(AAAAAAAG)_4$, and $d(TTTTTTTC)_4$ were purchased from Integrated DNA Technologies (Coralville, IA). Duplex poly(dA)·poly(dT) and single strand poly(dA) nucleotides were purchased from Amersham Pharmacia Biotech (Piscataway, NJ). Poly(dT) (lot 16H10741) was purchased from Sigma. Length of polymers in duplex poly(dA)·poly(dT) and single-stranded poly(dT) were greater than 500 nucleotides and poly(dA) was ~310 nucleotides in length. DNA lengths were confirmed by denaturing polyacrylamide gel electrophoresis. RNA duplex poly(rA)·poly(rU) and single strand poly(rA) and poly(rU) were purchased from Dharmacon. RNA oligonucleotides were deprotected according to protocols by the manufacturer. Coralyne chloride (lot

106C0362) was purchased from Sigma (St. Louis, MO). Polynucleotides and coralyne were used without further purification.

2.2.2. Sample Preparation

Oligonucleotides (32-mers and 16-mers) were purified on a 1 m G-25 Sephadex column. Fractions containing the purified oligonucleotides were pooled, lyophilized and the DNA was resuspended in dH_2O. Concentrations of oligonucleotides were determined by UV-Vis spectroscopy using the following extinction coefficients [72]: poly(dA)·poly(dT), ε_{260} = 12000 M^{-1} cm^{-1} per base pair; poly(rA)·poly(U), ε_{260} = 14280 M^{-1} cm^{-1}, 32-mer & 16-mer (dT), ε_{264} = 8520 M^{-1} cm^{-1} per base; poly, 32-mer & 16-mer (dA), ε_{257} = 8600 M^{-1} cm^{-1} per base; coralyne chloride, ε_{420} = 14500 M^{-1} cm^{-1}. Poly (rU), ε_{260} = 9350 M^{-1} cm^{-1}, was determined using the following reference [119]. All DNA samples were 55 μM per base, base pair, or base triplet, 115 mM NaCl, 13 mM Na-cacodylic buffer, pH 6.8; unless otherwise stated. Coralyne chloride concentrations are given in figure captions.

2.2.3. Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra were acquired on a JASCO J-720 CD spectropolarimeter equipped with an RTE-111 temperature control unit. Spectra were acquired using a 10 mm path length cell. CD cuvettes were purchased from Helma Cells. CD melting profiles were acquired by increasing the sample temperature at a rate of 0.8°C/min. The scan speed was 500 nm/min. Spectra are an average of 5 scans. Integration time was 1 sec.

2.2.4. UV-Vis Spectroscopy

UV-Vis absorbance measurements were performed using an HP 8453 UV-Vis diode array spectrophotometer instrument with an Agilent 89090A Peltier temperature control unit. Spectra were acquired using a 10 mm path length cell purchased from Starna Cells. Integration time was set at 0.5 sec.

2.3. RESULTS AND DISCUSSION

2.3.1. Complete Triplex Formation Requires Heating in the Presence of Coralyne

Our studies have found that complete triplex formation and coralyne intercalation by $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ is blocked by a kinetic trap at 4ºC. The pyrimidine triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ is not stable under the solution conditions of our study, even at 4°C. This is illustrated by the fact that a sample with a 1:2 molar ratio of $(dA)_{16}$ and $(dT)_{16}$ exhibits a CD spectrum that is virtually identical to that of the $(dA)_{16}\cdot(dT)_{16}$ duplex in a sample containing a 1:1 molar ratio of $(dA)_{16}$ and $(dT)_{16}$ (Figure 2.2A). Additionally, the CD melting profiles of these two samples exhibit the same single melting transition (i.e. $T_m^{2\rightarrow1}$) at 37ºC (Figure 2.3A). Thus, the only secondary structure present in the sample with a 1:2 molar ratio of $(dA)_{16}$ and $(dT)_{16}$ is the duplex $(dA)_{16}\cdot(dT)_{16}$, which coexists with an equal molar equivalent of single stranded $(dT)_{16}$. We will refer to a sample in this state as "$(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$", as opposed to $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$, to emphasize the absence of triplex secondary structure in a sample that contains a 1:2 molar ratio of $(dA)_{16}$ and $(dT)_{16}$ strands. The absence of triplex secondary structure in the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample is somewhat particular to the conditions of our study, as the addition of divalent cations can facilitate triplex formation by oligonucleotides of this length and

sequence at 4ºC [120]. However, in order to be consistent with our past studies [88], and given our specific aim to explore nucleic acid stabilization by intercalation, we have used solution conditions in which the triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ is unstable in the absence of coralyne intercalation.

Coralyne has previously been shown to intercalate triplex DNA and thereby enhance the thermal stability of the triplex secondary structure [115]. The addition of coralyne to a $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample at 4°C results in an appreciable change in the sample CD spectrum (Figure 2.2B). However, *complete* triplex formation and coralyne intercalation does not appear to occur upon coralyne addition, as a triplex intercalated by coralyne produces small positive CD bands at ~340 and ~440 nm [88], which is not observed. The CD bands that appear upon the addition of coralyne to the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample at 4°C are potentially due to a non-intercalative mode of coralyne binding to duplex secondary structure in the sample. The persistence of duplex secondary structure that interacts with coralyne is supported by the observation of similar (but more intense) CD bands at around 295 nm for a sample of duplex $(dA)_{16}\cdot(dT)_{16}$ with coralyne at the same temperature (*vide infra*). In any case, the shape of the CD spectrum of the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample at 4°C with added coralyne is not typical of an intercalated triplex, and this indicates that coralyne does *not* cause the complete and spontaneous formation of a triplex in this sample when it is maintained at 4°C. This is surprising given the fact that coralyne is well known to greatly increase the melting temperature of A·T·T triplexes [110].

Heating the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample with added coralyne to 30°C causes the CD spectrum of this sample to adopt spectral features that are typical of a DNA triplex

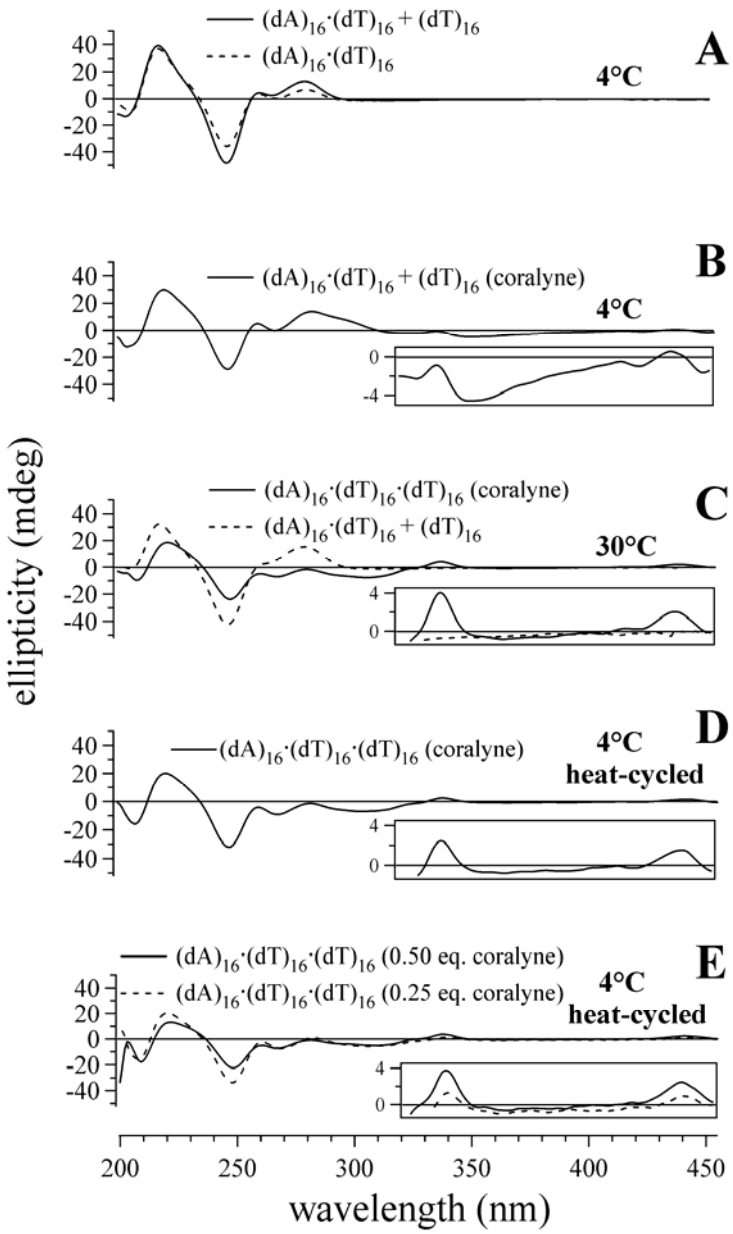


Figure 2.2. Circular dichroism spectra as a function of wavelength illustrate coralyne-mediated triplex formation. CD spectra of triplex and duplex 16-mer DNA samples demonstrating that complete triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ formation requires heating of the DNA sample in the presence of coralyne. (A) Spectra of duplex $(dA)_{16}\cdot(dT)_{16}$ and $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ (unstable triplex) samples at 4°C. (B) Spectrum of $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ in the presence of coralyne at 4°C (C) Spectrum of $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ at 30°C without coralyne, and triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ spectrum with coralyne at 30°C. (D) CD spectrum acquired at 4°C of $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ after being heated to 75°C with coralyne. (E) Spectra of triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ at 4°C after heat cycling with 0.50 molar equivalents and 0.25 molar eq. of coralyne, respectively. Coralyne concentration was 0.25 molar equivalents per base pair and base triplet, except as stated in E.

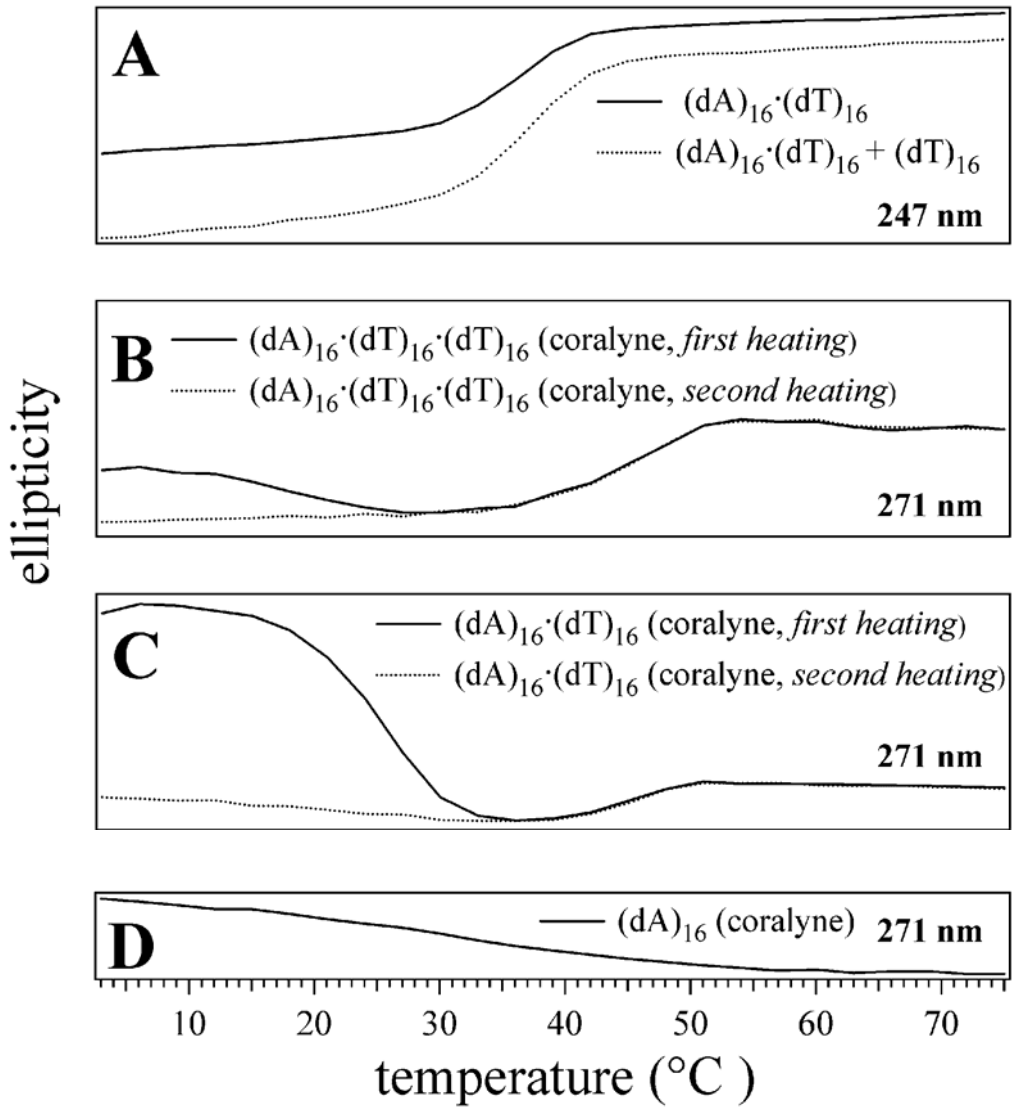


Figure 2.3. CD melting profiles for 16-mer DNA samples at selected wavelengths showing structural transitions in DNA samples. (A) Duplex $(dA)_{16}\cdot(dT)_{16}$ and $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ melting curves exhibit a transition ($T_m^{2\rightarrow1}$) at 37°C. (B) First and second heating of triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ sample after the addition of coralyne indicates triplex melting ($T_m^{3\rightarrow1}$) at 46°C. The first heating of this sample after the addition of coralyne also shows a transition centered at ~15°C, which is assigned to the reorganization of DNA strands form partial duplex and partial triplex to complete triplex. (C) First and second heating of duplex $(dA)_{16}\cdot(dT)_{16}$ after the addition of coralyne; the duplex disproportionation transition is observed at ~23°C during the first heating and triplex melting ($T_m^{3\rightarrow1}$) is observed at 46°C in both the first and second heating of the sample. (D) $(dA)_{16}$ with coralyne exhibits the melting transition of the $(dA)_{16}$ self-structure at ~25°C. Vertical ellipticity scale is the same for all samples. Coralyne concentrations were 0.25 molar equivalents per base, base pair or base triplet.

with intercalated coralyne, including the appearance of small positive CD bands at ~340 and ~440 nm (Figure 2.2C). The CD spectrum of the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample at 30°C *without* coralyne, on the other hand, remains similar to that of duplex $(dA)_{16}\cdot(dT)_{16}$ (Figures 2.2A & 2.2C).

Thus, heating of the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample with coralyne to 30°C actually promotes the formation of triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$. The CD heating profile for triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ with added coralyne reveals that the intercalated triplex melts at 46°C into single strands in a single transition ($T_m^{3\rightarrow1}$) (Figure 2.3B). The CD spectrum of triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ with coralyne at 4°C *after* being heat cycled above 30°C is dramatically different from the CD spectrum of the same sample at 4°C *prior* to being heated (Figures 2.2B & 2.2D). This demonstrates that the transition to a sample of complete coralyne-intercalated triplex is not reversed upon cooling. This is also illustrated by the fact that the CD melting profile for the first heating of the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample with coralyne has a transition at ~15°C, assigned to DNA strand reorganization, that is absent during the second heating of the sample (Figure 2.3B).

UV-Vis spectrophotometry provides additional evidence that heat cycling of the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample changes the nature of coralyne-DNA interactions (Figure 2.4). Coralyne has two absorption bands with local maxima at ~420 and ~440 nm that appear with DNA intercalation [114]. A spectrum acquired before heating is indicative of a mixed, but not fully intercalative, mode of coralyne interaction with $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$. After heat cycling, at 4°C the intensity of the coralyne band at 440 nm for the same sample has increased relative to the 420 nm absorption band (Figure 3). This is

consistent with an increase in the level of coralyne intercalation into the $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ triplex after heating [114].

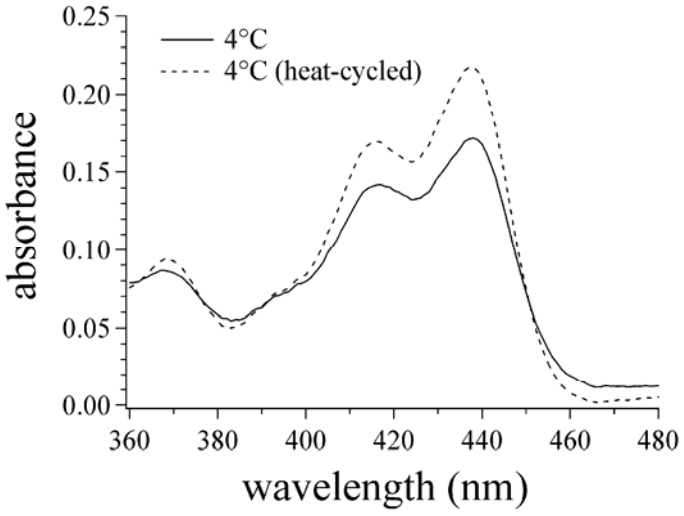


Figure 2.4. UV-Vis spectroscopy illustrates coralyne binding to a triplex. The 360-480 nm region of the UV-Vis absorption spectra of coralyne in the presence of triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$. Spectra at 4°C before & after heating illustrate the change in binding of coralyne to the 16-mer triplex that is promoted by heat cycling above 36°C. DNA conc. was 142 μM per base triplet and coralyne conc. was 0.25 molar eq. per base triplet.

The coralyne-intercalated (heat cycled) triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ samples discussed thus far were intercalated only to a level of one coralyne per four base triplets, i.e. 0.25 equivalents of coralyne per base triplet. Coralyne was initially added to a $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample to a concentration that corresponded to 0.50 molar equivalents of coralyne per (potential) base triplet (i.e. 27.5 μM). This ratio of coralyne would allow intercalation of the triplex to the level of one coralyne per every other base triplet, the maximum level allowed under the nearest neighbor exclusion principle. However, the addition of coralyne to this concentration in a $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample at 4°C immediately resulted in clouding of the sample and subsequent precipitation of

coralyne. It was found that coralyne could be added to 0.25 molar equivalents of coralyne per base triplet at 4°C without precipitation. Once this sample was heated-cycled up to 75°C and back to 4°C, the coralyne concentration could be increased to 0.5 molecules per base triplet without sample clouding or coralyne precipitation. Heat cycling of the sample with the full 0.5 molar equivalents of coralyne then resulted in an increase in coralyne intercalation, based upon the intensity of CD bands at ~340 and ~440 nm that approximately doubled in comparison with the sample that contained only 0.25 molar equivalent of coralyne (Figure 2.2E). Thus, it appears that upon heating, coralyne causes a rearrangement of the DNA strands in such a manner (i.e. triplex formation) that provides additional sites for coralyne intercalation.

The ability for small molecule intercalation to stabilize triplex structures has traditionally been demonstrated by measuring a shift in the T_m of a triplex in the presence of the small molecule [103]. However, the data presented here illustrates that if a triplex is not stable at the temperature at which the small molecule is added to the solution, heating may be required to facilitate complete triplex formation and intercalation. This would likely have gone unnoticed in past studies where only sample absorption at 260 nm was monitored as a function of temperature. Nevertheless, our results indicate that the ability for a small molecule to stabilize a triplex could depend upon the initial state of the sample, as the duplex structure can apparently act as a kinetic barrier to triplex formation. This may have significant implications regarding the proposed use of intercalators to stabilize triplex structures under conditions where the triplex is not stable in the absence of intercalation.

2.3.2. Coralyne Causes Disproportionation of Duplex $(dA)_{16}\cdot(dT)_{16}$ at 30°C

The CD spectrum of duplex $(dA)_{16}\cdot(dT)_{16}$ at 4°C changes dramatically upon the addition of 0.25 molar equivalents of coralyne per DNA base pair (Figure 2.5A). These changes include the appearance of a substantial positive CD band (or bands) near 280 nm and negative CD bands near 340 nm. As noted above, similar CD bands near 280 and 340 nm were also observed upon the addition of coralyne to the $(dA)_{16}\cdot(dT)_{16}$ + $(dT)_{16}$ sample at 4°C. However, these CD bands are more pronounced in the duplex $(dA)_{16}\cdot(dT)_{16}$ sample (Figure 2.5A). Coralyne has been reported to interact with DNA by both intercalative and non-intercalative binding [114]. These CD bands may result from the binding of coralyne in the minor groove of $(dA)_{16}\cdot(dT)_{16}$, as AT-rich regions of DNA duplexes are known to bind a variety of small molecules in the minor groove [76, 121, 122]. Additionally, cationic cyanine dyes that bind in the minor groove of duplex DNA have previously been shown to cause the appearance of positive CD bands in DNA samples [123].

Several experimental observations indicate that upon heating, duplex $(dA)_{16}\cdot(dT)_{16}$ in the presence of coralyne undergoes disproportionation into 0.50 molar equivalents of triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ and 0.50 molar equivalents of $(dA)_{16}$. These observations include: Heating of the $(dA)_{16}\cdot(dT)_{16}$ sample with added coralyne to 36°C dramatically reduces the magnitude of the duplex-specific coralyne CD bands at 280 and 340 nm (Figure 2.5B); The shape of the DNA bands between 200 and 260 nm change with heating to more closely resemble those of a DNA triplex (Figure 2.5B); The CD spectrum of the $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne at 36°C shows small positive bands at ~340 and ~440 nm that are indicative of coralyne intercalation.

Additionally, the melting profile of the $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne added exhibits a transition at 46°C (Figure 2.3C), which is the same temperature at which the coralyne-intercalated triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ melts (Figure 2.3B). The magnitude of this transition in the CD melting profile of the duplex sample with added coralyne is one half the transition in the triplex sample with added coralyne (Figures 2.3B & 2.3C), which is also consistent with the $(dA)_{16}\cdot(dT)_{16}$ sample being disproportioned by coralyne into 0.5 molar equivalents of triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ and 0.5 molar equivalents of $(dA)_{16}$. Thus, the 46°C transition in the $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne can be assigned to the melting of coralyne-intercalated triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ in the disproportioned sample.

The process of duplex disproportionation by coralyne in the $(dA)_{16}\cdot(dT)_{16}$ sample during the *first* sample heating is indicated by a broad transition that is centered at ~23°C (Figure 2.3C). During the second heating of the same sample, this broad transition is absent (Figure 2.3C). This indicates that the intercalated triplex and $(dA)_{16}$ of a coralyne-disproportioned duplex sample do not immediately revert to the duplex state when the sample is returned to 4°C. This lack of reversion from disproportionation is also supported by the fact that the CD spectrum of the coralyne-disproportioned $(dA)_{16}\cdot(dT)_{16}$ at 4°C after heat cycling, to 75°C and back to 4°C, is radically different from the CD spectrum of the sample prior to heating (Figures 2.5A & 2.5C). Furthermore, there is an excellent match between the CD spectrum of coralyne-disproportioned $(dA)_{16}\cdot(dT)_{16}$ sample and a composite CD spectrum generated by the summation of a CD spectrum of coralyne-intercalated triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ and the CD spectrum of $(dA)_{16}$ in the presence of coralyne (Figure 2.5C).

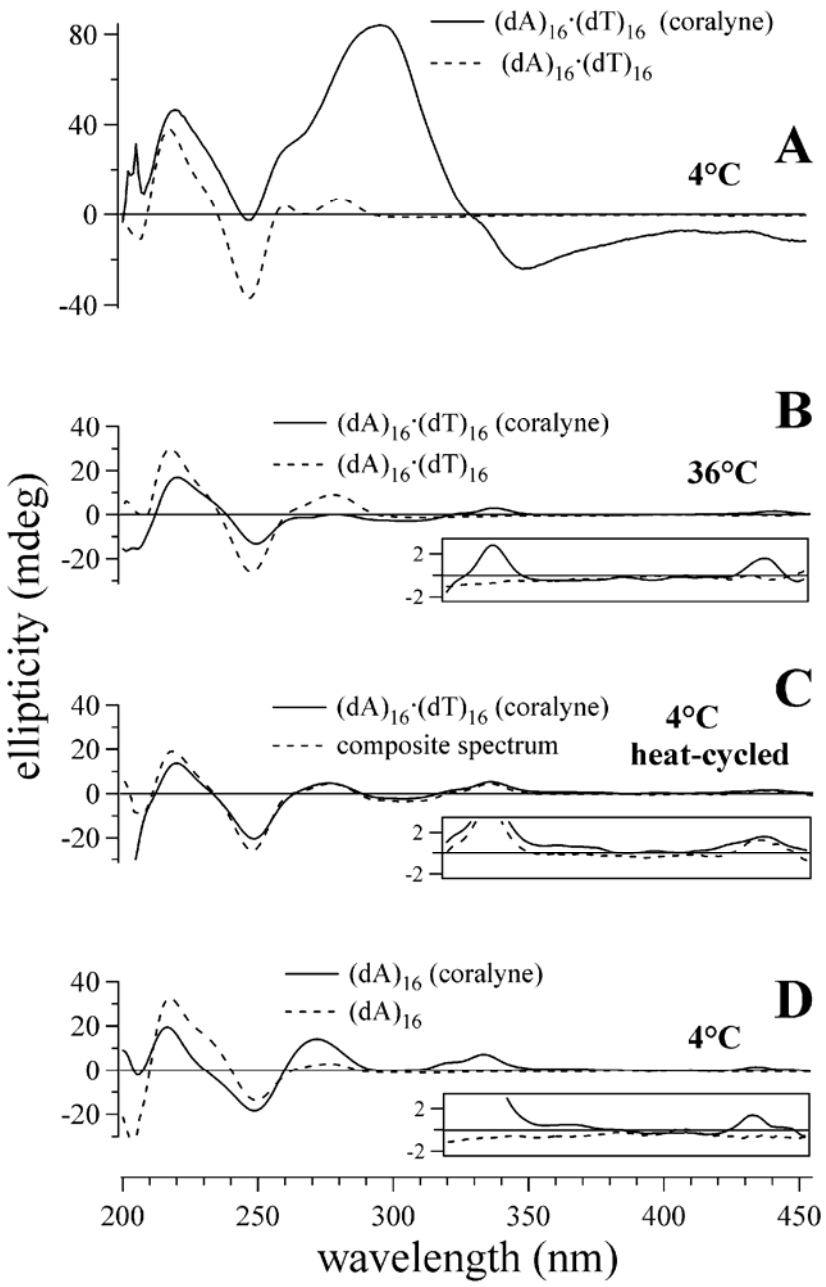


Figure 2.5. Circular dichroism spectra of $(dA)_{16}\cdot(dT)_{16}$ samples illustrating duplex disproportionation. (A) Duplex $(dA)_{16}\cdot(dT)_{16}$ sample with and without added coralyne at 4°C prior to heating. (B) Spectra of duplex $(dA)_{16}\cdot(dT)_{16}$ sample with and without coralyne at 36°C. (C) Heat-cycled spectrum of $(dA)_{16}\cdot(dT)_{16}$ disproportioned sample in the presence of coralyne at 4°C, and composite spectrum (sum of spectra acquired at 4°C: (0.50 × triplex $(dA)_{16}\cdot(dT)_{16}\cdot(dT)_{16}$ with 0.25 molar equivalents of coralyne) + (0.50 × single stranded $(dA)_{16}$ with 0.25 molar equivalents of coralyne). (D) Spectra of $(dA)_{16}$ sample with and without coralyne at 4°C. Coralyne concentration was 0.25 molar equivalents per base, base pair or base triplet.

Our previous investigations revealed that the poly(dA) liberated as a result of duplex disproportionation adopts a self-structure with A·A base pairs that is completely dependent on coralyne intercalation for stability [88]. Here, we have shown that that much shorter homo-dA strands also form the $(dA)_n$ self-structure in the presence of coralyne [86]. In Figure 2.5D, the CD spectra are presented for $(dA)_{16}$ at 4°C in the presence and absence of coralyne. The addition of coralyne to the $(dA)_{16}$ sample leads to the appearance of a significant CD band at ~340 nm that indicates coralyne binding. The CD spectrum of $(dA)_{16}$ in the presence of coralyne between 220 and 270 nm also differs significantly from $(dA)_{16}$ in the absence of coralyne, indicating a significant change in the secondary structure of $(dA)_{16}$ upon coralyne binding. The CD melting profile for $(dA)_{16}$ with coralyne shows a relatively broad melting transition centered at ~25°C (Figure 2.3D).

2.3.3. Coralyne-Induced $(dA)_{16}$·$(dT)_{16}$ Disproportionation is Reversible

We have found that after heat cycling, the CD spectrum of coralyne-disproportioned $(dA)_{16}$·$(dT)_{16}$ changes over the course of hours if the sample is maintained at 4°C. In Figure 2.6, CD spectra are presented that illustrate this temporal change. For the particular experiment shown, 0.25 molar equivalents of coralyne were added to a $(dA)_{16}$·$(dT)_{16}$ sample at 4°C, the sample was then heated to 75°C and cooled back to 4°C. It is clear that over time the disproportioned duplex sample reverts back to the duplex state, as the CD spectrum after 20 hours is virtually identical to the spectrum of the sample acquired prior to heating and immediately following the addition of coralyne at 4°C (Figure 2.6A). CD spectra of the disproportioned sample acquired over

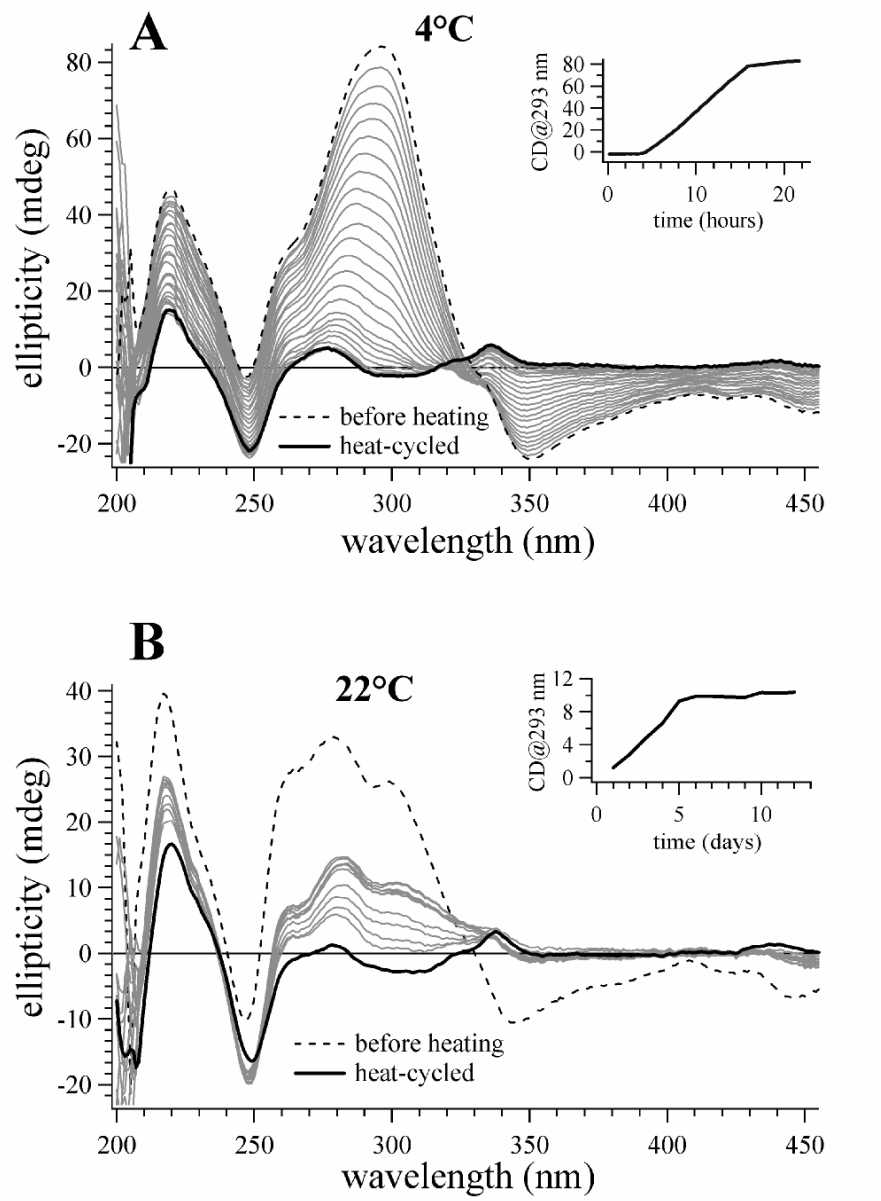


Figure 2.6. CD spectra of a $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne illustrate equilibrium at 4°C and 22°C. (A) Spectra of the $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne at 4°C. Dashed line is the spectrum acquired immediately after the addition of coralyne at 4°C, and before heat cycling. Solid black line is the spectrum acquired immediately after the sample was heated to 75°C and cooled back to 4°C. Spectra drawn in gray were acquired at selected time intervals after the sample was heat cycled. After heat cycling, the sample was constantly maintained at 4°C. (Insert to A) A plot of the CD signal at 293 nm as a function of time after sample heat cycling for the sample maintained at 4°C. (B) Spectra of the $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne at 22°C (same nomenclature as A). (Insert to B) A plot of the best-fit exponential function of the CD signal at 293 nm for the $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne at 22°C as a function of time after heat cycling. Coralyne concentration was 0.25 molar equivalents per base pair for both samples.

the course of reversion at 4°C are well represented as the linear sum of the CD spectrum of the sample before heating (100% duplex) and the CD spectrum immediately following heat cycling (i.e. 100% disproportioned). Thus, the reversion of the sample from the disproportioned state can be modeled as a two state system. A plot of the magnitude of the CD signal at 293 nm as a function of time illustrates that duplex reversion is sigmoidal with time and is complete after approximately 16 hours (Figure 2.6A, inset).

The fact that coralyne-disproportioned $(dA)_{16}\cdot(dT)_{16}$ slowly reverts to the duplex state at 4°C indicates that duplex $(dA)_{16}\cdot(dT)_{16}$ is the thermodynamically favored state at 4°C, with respect to the disproportioned duplex. This explains why a $(dA)_{16}\cdot(dT)_{16}$ duplex sample does not spontaneously undergo disproportionation at 4°C when coralyne is added. Thus, the duplex at 4°C is not a kinetic barrier to disproportionation, which might be assumed because sample heating is required to initiate disproportionation. In contrast, reversion of a coralyne-disproportioned $(dA)_{16}\cdot(dT)_{16}$ sample back to the duplex state at 4°C is apparently slow because the disproportioned state is a kinetic barrier to reversion at 4°C.

2.3.4. Three Distinct DNA Secondary Structures Co-Exist in Equilibrium at 22°C

Our observation that the coralyne-disproportioned state of $(dA)_{16}\cdot(dT)_{16}$ is thermodynamically favored at 30°C, whereas the duplex state is favored at 4°C, immediately suggested that for some range of temperature, duplex $(dA)_{16}\cdot(dT)_{16}$ and coralyne-disproportioned $(dA)_{16}\cdot(dT)_{16}$ will coexist in equilibrium. To directly investigate this possibility, we studied the equilibrium state of $(dA)_{16}\cdot(dT)_{16}$ sample in the presence of coralyne at 22°C (room temperature). For this investigation, 0.25 molar equivalents of

coralyne were added to a duplex $(dA)_{16}\cdot(dT)_{16}$ sample at 22°C and the CD spectrum was acquired immediately (Figure 2.6B). The sample was then heat cycled between 75°C and 4°C to ensure complete disproportionation. The disproportioned sample was then moved to 22°C where it was maintained for several days, during which time CD spectra were collected at 22°C on a regular basis. The CD spectrum of the sample changed with time, as illustrated by a graph of the CD signal at 293 nm, until achieving equilibrium after approximately 5 days (Figure 2.6B, inset). The CD spectrum of the sample at equilibrium is intermediate to the CD spectrum of the sample at 22°C before heating to disproportionation and the CD spectrum acquired at 22°C immediately after disproportionation. Based upon the least-squares best-fit of the equilibrium spectrum as a weighted sum of the spectra before and after disproportionation, it appears that the equilibrium state of the $(dA)_{16}\cdot(dT)_{16}$ sample with coralyne at 22°C is 40% duplex and 60% disproportioned duplex (i.e. triplex and $(dA)_{16}$ self-structure) (Figure 2.7).

Previous studies have shown that increasing the temperature of a $(dA)_n\cdot(dT)_n$ duplex sample with divalent cations or high monovalent cation concentration can cause duplex disproportionation in the absence of intercalation [104, 124, 125]. However, without intercalation, duplex disproportionation is typically not complete before the sample reaches a temperature at which all secondary structures melted into single strands, and disproportionation is readily reversed upon a decrease in sample temperature [104, 124, 125]. Our results indicate that coralyne intercalation is working synergistically with temperature to drive duplex disproportionation, for both intercalation and increased temperature (i.e. above ~30°C) appear to be necessary for our duplex DNA sample to achieve complete disproportionation.

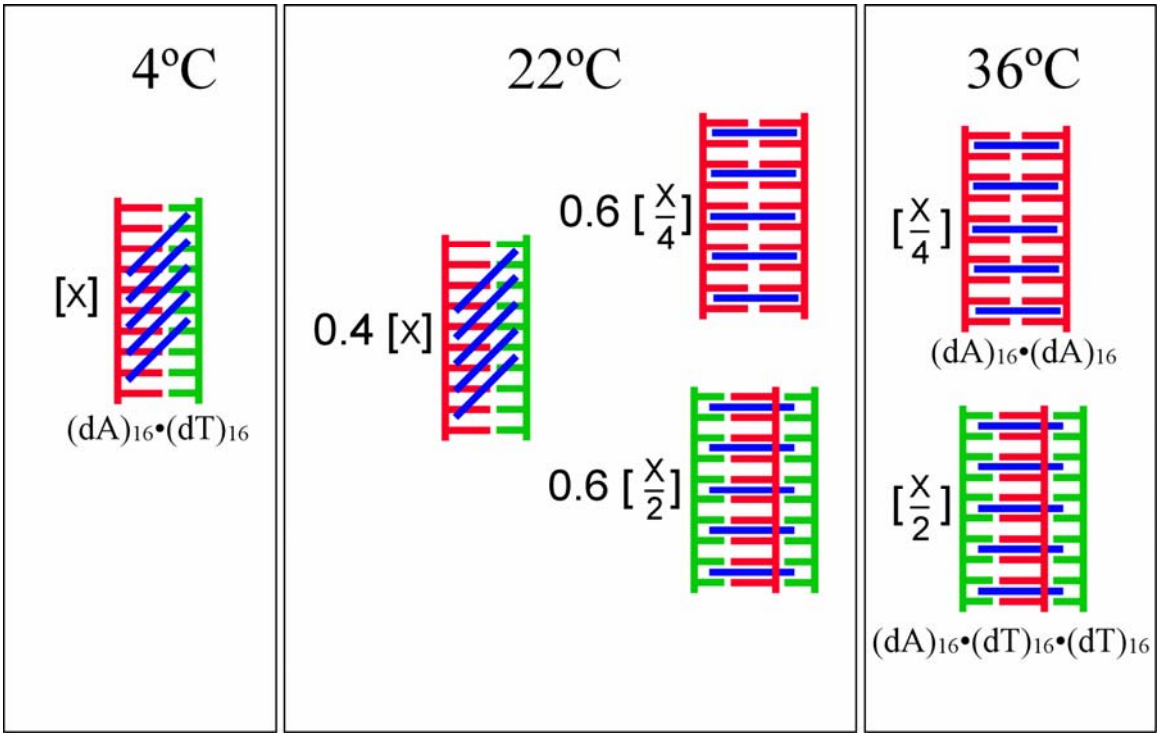


Figure 2.7. Schematic representation of the equilibrium secondary structure distribution at three temperatures for a sample containing an equal ratio of $(dA)_{16}$ and $(dT)_{16}$ in the presence of 0.25 molar equivalents of coralyne per base pair. The possibility of alternative mode of coralyne binding to the $(dA)_{16}$·$(dT)_{16}$ duplex is depicted by the blue diagonal lines

2.3.5. $(dA)_{32}\cdot(dT)_{32}$ and Poly(dA)·poly(dT) Can Require Weeks to Reach Equilibrium

We previously reported that coralyne causes the *irreversible* disproportionation of duplex poly(dA)·poly(dT) based upon the observation that the spectrum of a coralyne-disproportioned poly(dA)·poly(dT) sample was stable at 4°C for more than one day [88]. We also observed that duplex poly(dA)·poly(dT) does not undergo disproportionation for a matter of months if coralyne is added at 4°C and the sample is continually maintained at 4°C. Combined, these observations lead us to conclude that duplex poly(dA)·poly(dT) does not spontaneously disproportion when coralyne is added at 4°C because the duplex structure (with associated coralyne) acts as a kinetic barrier to disproportionation, and that heating above 30°C is necessary to over come this barrier. However, the results presented above with $(dA)_{16}\cdot(dT)_{16}$ suggest a different assignment of which DNA secondary structures of poly(dA)·poly(dT) in the presence of coralyne at 4°C are thermodynamically favored and which are kinetic traps.

To examine the effect of DNA length on the propensity for a disproportioned sample to revert to the duplex state at 4°C, we repeated the same experiment described above for monitoring $(dA)_{16}\cdot(dT)_{16}$ reversion, except with the 32-mer duplex $(dA)_{32}\cdot(dT)_{32}$. The CD spectra of coralyne-disproportioned $(dA)_{32}\cdot(dT)_{32}$ also changes over time at 4°C (Figure 2.8A). However, there are two significant differences between 32-mer and 16-mer duplex reversion. First, coralyne-disproportioned $(dA)_{32}\cdot(dT)_{32}$ requires several days to reach equilibrium at 4°C, rather than several hours, with an exponential decay time constant of almost 7 days (Figure 2.8A, inset). Secondly, the equilibrium state that the 32-mer sample at 4°C is partially disproportioned state, much

like that reached by the 16-mer sample at 22°C. Thus, doubling oligonucleotide length from 16 to 32 actually shifts the secondary structure equilibrium at 4°C towards the disproportioned state.

The reversion of a coralyne-disproportioned poly(dA)·poly(dT) sample at 4°C was investigated as well. In an experiment similar to those described above, 0.25 molar equivalents of coralyne were added to a poly(dA)·poly(dT) sample at 4°C. This sample was then heated to 95°C, above the coralyne-intercalated poly(dA)·poly(dT)·poly(dT) melting temperature (i.e. 87°C), and then cooled back to 4°C. The disproportioned sample was maintained and monitored by CD at 4°C for several months (Figure 2.8B). CD spectra indicate that the secondary structures within a coralyne-disproportioned poly(dA)·poly(dT) sample at 4°C do change over time, however, the equilibrium state of this sample is more difficult to interpret than that of the 16-mer or the 32-mer samples. The approach to equilibrium by the polynucleotide sample is best fit by a double exponential, with a fast time constant of 1.4 days and a slow time constant of 28 days (Figure 2.8B, insert). The equilibrium state that is finally reached by this sample does not appear to be intermediate to that of the same sample before and after heating (Figure 2.8B). Thus, changing DNA length from 32 to ~300 nucleotides (i.e. poly(dA), poly(dT)) can change the secondary structures that are favored by a $(dA)_n \cdot (dT)_n \cdot (dT)_n$ sample in the presence of coralyne at 4°C.

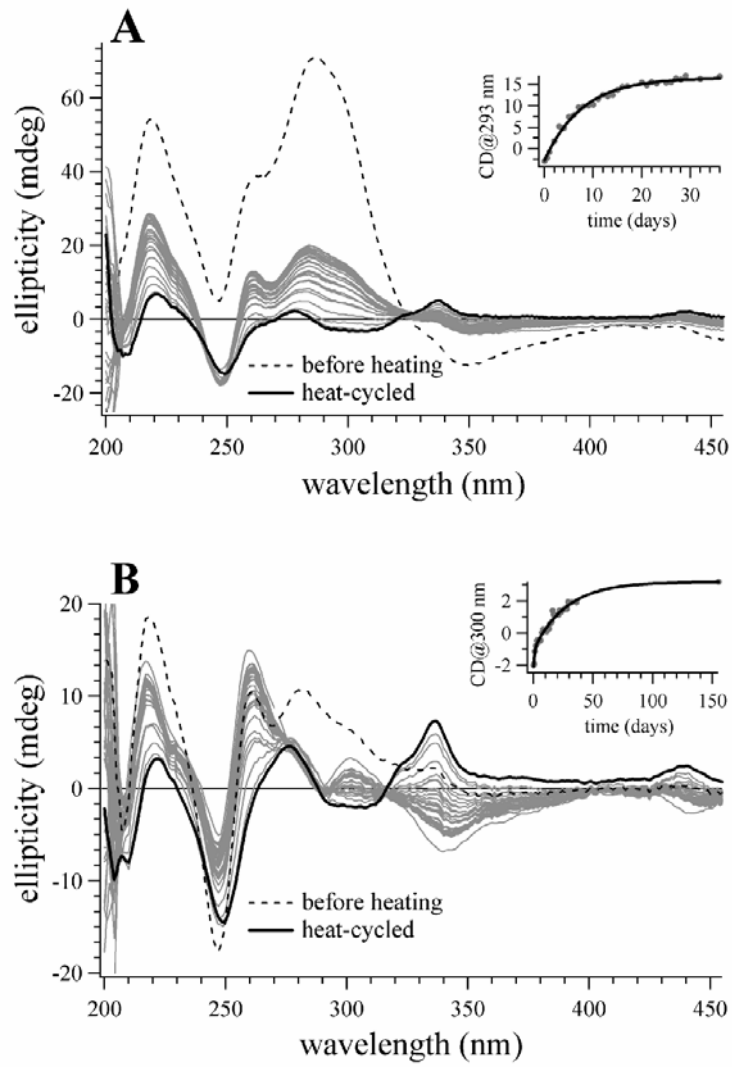


Figure 2.8. CD spectra of $(dA)_{32}\cdot(dT)_{32}$ and poly(dA)·poly(dT) samples with coralyne that illustrate the approach of these sample to equilibrium at 4°C after being heat cycled into the disproportioned state. (A) Spectra of the $(dA)_{32}\cdot(dT)_{32}$ sample with coralyne at 4°C. Dashed line is the spectrum acquired immediately after the addition of coralyne at 4°C, and before sample heat cycling. Solid black line is the spectrum acquired immediately after the sample was heated to 75°C and cooled back to 4°C. Spectra drawn in gray were acquired at selected time intervals after the sample was heat cycled. After heat cycling, the sample was constantly maintained at 4°C. (Insert to A) A plot of the best-fit exponential function of the CD signal at 293 nm for the $(dA)_{32}\cdot(dT)_{32}$ sample with coralyne at 4°C as a function of time after sample heat cycling. (B) Spectra of the poly(dA)·poly(dT) sample with coralyne at 4°C. (Insert to A) A plot of the best-fit double-exponential function of the CD signal at 293 for the poly(dA)·poly(dT) sample with coralyne at 4°C nm as a function of time after heat cycling. Coralyne concentration was 0.25 molar equivalents per base pair for both samples.

2.3.6. Effect of G•C Base Pairs on Coralyne-Driven Disproportionation

We have observed that coralyne causes disproportionation of a duplex $(dA)_{32}\cdot(dT)_{32}$ sample into a triplex $(dA)_{32}\cdot(dT)_{32}\cdot(dT)_{32}$ and $(dA)_{32}$ self-structure. We wanted to investigate the effects of having G•C pairs in a $(dA)_n\bullet(dT)_n$ duplex structure which is expected to form a C•G•C base triplets upon disproportionation and triplex formation (Figure 2.9). For this study, we have taken 2 different sets of 32-mer sequences and inserted G and C residue in separate strands which are expected to form G•C pairs in a duplex and C•G•C triplets in a DNA triplex structure. Coralyne specifically intercalates T•A•T triplets. Therefore, we expected the presence of C^+•G•C triplex to have a deleterious effect on triplex formation.

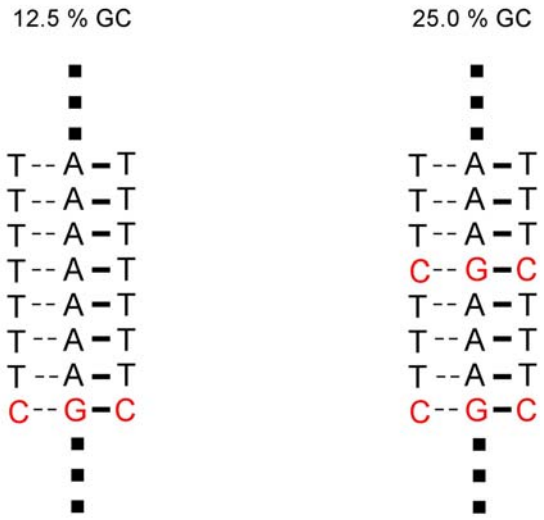


Figure 2.9. Schematic representation of the 32-mer DNA triplexes containing C•G•C base triplets. Triplex labeled as 25.0% GC (left) contains C•G•C triplet every 4th base step whereas the structure labeled 12.5% GC (right) contains C•G•C triplets every 8th base step.

Our results show that the triplex structure (labeled 12.5% GC in the Figure 2.9) is not stable under the solution condition of our study. The CD melting profiles of the

duplex sample (1:1 molar ratio of $d(AAAGAAAG)_4$ and $d(TTTCTTTC)_4$) and the triplex sample (1:2 molar ratio of $d(AAAGAAAG)_4$ and $d(TTTCTTTC)_4$) exhibit the same single melting transition (i.e. $T_m^{2\rightarrow1}$) at 52ºC (Figure 2.10). Thus, the only secondary structure present in the sample with a 1:2 molar ratio of $d(AAAGAAAG)_4$ and $d(TTTCTTTC)_4$) is the duplex $d(AAAGAAAG)_4$•$d(TTTCTTTC)_4$), which coexists with an equal molar equivalent of single stranded $(dTTTCTTTC)_4$.

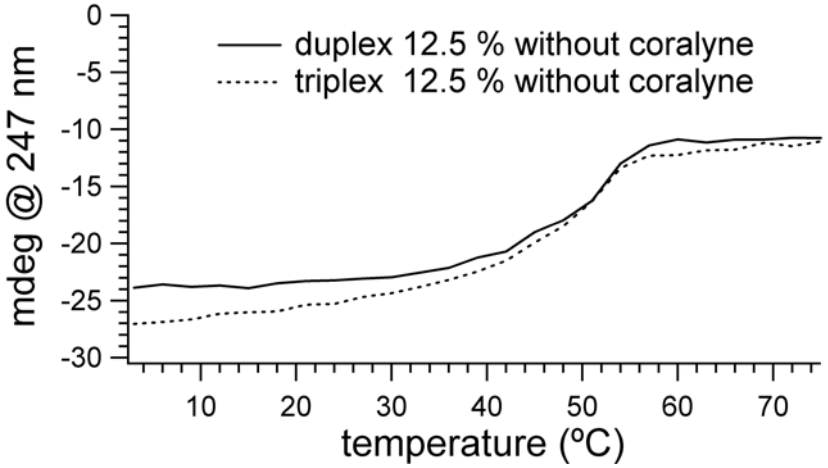


Figure 2.10. Circular dichroism melting analysis of a 32-mer duplex containing G•C base pairs. CD melting profiles for 12.5% GC duplex $d(AAAGAAAG)_4$•$d(TTTCTTTC)_4$ (solid) and 12.5% GC triplex $d(AAAGAAAG)_4$•$d(TTTCTTTC)_4$•$d(TTTCTTTC)_4$ (dashed) samples without coralyne at 247 nm exhibit melting transition centered at 52ºC. Solution conditions are indicated in the experimental section.

We find that upon heating in the presence of coralyne, the duplex $d(AAAAAAAG)_4$•$d(TTTTTTTC)_4$ sample undergoes changes in the structure which are indicative of a triplex $d(AAAAAAAG)_4$•$d(TTTTTTTC)_4$•$d(TTTTTTTC)_4$ with intercalated coralyne (Figure 2.11) [86, 88]. The CD spectrum of the duplex sample with coralyne at 51°C shows small positive bands at ~340 and ~440 nm which indicate coralyne intercalation into a triplex structure.

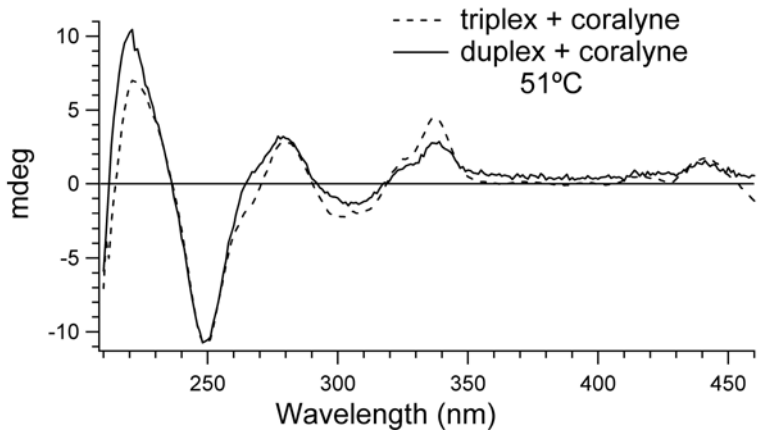


Figure 2.11. CD spectroscopy illustrating coralyne interactions with 32-mer structures. Spectra of a 12.5% GC triplex (dashed) and duplex (solid) sample in the presence of coralyne at 51ºC show coralyne intercalation bands centered at ~ 340 nm and ~ 440 nm. Coralyne concentration was 0.50 molar equivalents per base pair and base triplet. Solution conditions were 55 μM in base pair or base triplet, 115 mM Na^+, 13 mM cacodylate buffer (pH 6.8). Spectra collected at 51ºC.

CD melting profile of a duplex $d(AAAAAAAG)_4$•$d(TTTTTTTC)_4$ with coralyne indicates two different structural transitions (Figure 2.12). Large transition centered at ~ 30ºC could be potentially due to duplex $d(AAAAAAAG)_4$•$d(TTTTTTTC)_4$ disproportionation into 0.5 molar equivalents of triplex $d(AAAAAAAG)_4$•$d(TTTTTTTC)_4$•$d(TTTTTTTC)_4$ and 0.5 molar equivalents of $d(AAAAAAAG)_4$. We have observed a similar type of structural transition for a $(dA)_{16}$•$(dT)_{16}$ duplex and a $(dA)_{32}$•$(dT)_{32}$ sample in our recent work [86]. We observe that the triplex melting in the presence of coralyne occurs at ~ 60ºC whereas the stability of the duplex structure is ~ 52ºC (Figure 2.11). Addition of coralyne to a sample containing $d(AAAAAAAG)_4$•$d(TTTTTTTC)_4$ + $d(TTTTTTTC)_4$ yields a melting profile that is very similar to that of a duplex sample with coralyne. However, the magnitudes of the first and second transitions in a triplex sample are significantly larger than the corresponding transitions in a duplex sample. We observe that the disproportioned sample containing

C•G•C triplets is not very stable upon sample cooling. It may altogether be possible that some or all molar equivalents of the duplex structure within the triplex sample undergo disproportionation.

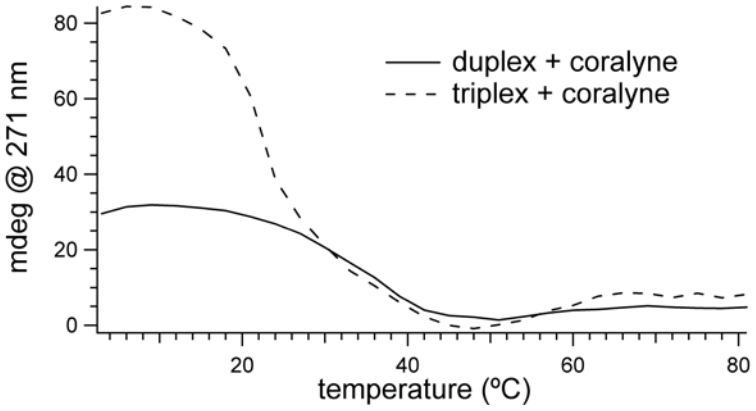


Figure 2.12. CD melting profile analysis of 32-mer structures containing G•C bp. Heating a 12.5% GC duplex in the presence of coralyne shows a duplex disproportionation transition at ~ 30ºC. Another transition at 60ºC is suggestive of triplex melting. A 12.5% GC triplex also shows similar transitions as that of the duplex sample. Coralyne concentration was 0.50 molar eq. per base pair and base triplet.

2.3.7. Coralyne Binds to an RNA Triplex and Single Stranded Poly(rA)

Previous work in our lab has shown that coralyne causes disproportionation of a duplex Poly(dA)•Poly(dT) sample into 0.5 molar equivalents of a triplex (Poly dA)•(Poly $dT)_2$ and 0.5 molar equivalents of a coralyne-intercalated Poly(dA) self structure [88]. We wanted to explore whether 1) coralyne can bind RNA duplex and triplex structures and 2) coralyne is able to drive structural transitions by stabilizing nucleic acid structures which are otherwise not stable in the absence of coralyne. Coralyne was added up to a concentration of 0.50 molar equivalents per base triplet to a (Poly rA)•(Poly $rU)_2$ sample. Spectra at 35ºC during the first heating show the appearance of a small positive band at ~

440 nm. During the 2nd heating, 2 positive bands at ~ 340 nm and ~ 440 nm are observed which are indicative of a triplex sample with intercalated coralyne [88].

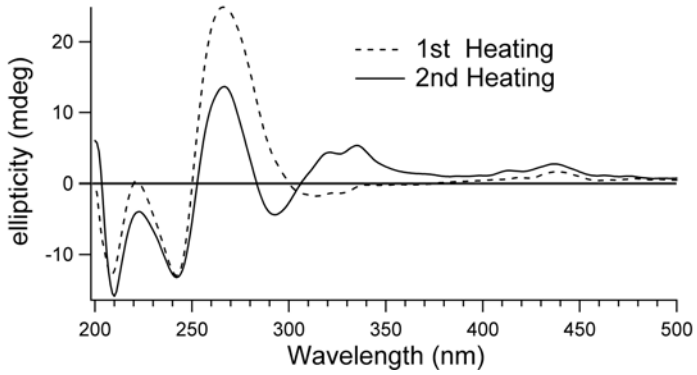


Figure 2.13. Coralyne interactions with RNA triplexes shown by CD. Spectra at 35ºC from 1st heating (dashed) and 2nd heating (solid) for a (Poly rA)•(Poly rU)$_2$ triplex with coralyne show significant changes in the structure upon coralyne addition. Coralyne concentration was 0.50 molar equivalents per base triplet.

Addition of coralyne to a duplex (Poly rU)•(Poly rA) sample also leads to the appearance of new bands in the CD spectra which maybe suggestive of coralyne-mediated disproportionation of a duplex (Poly rU)•(Poly rA) strand into a triplex (Poly rA)•(Poly rU)$_2$ structure (Figure 2.14).

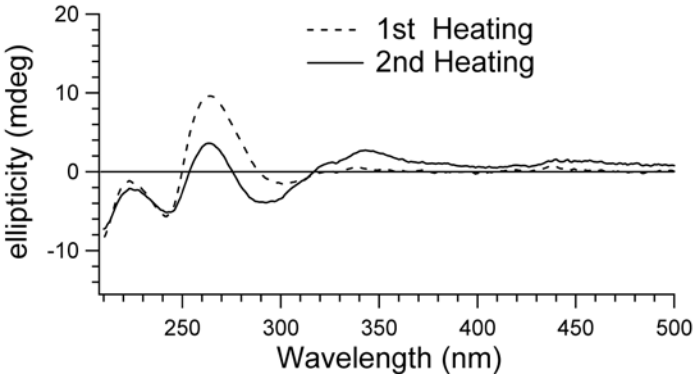


Figure 2.14. Coralyne interactions with RNA duplex structure depicted by CD. Spectra at 35ºC for 1st (dashed) and 2nd (solid) heating of a (Poly rU)•(Poly rA) duplex are shown. Changes in CD occur during the second heating of the sample with the appearance of a positive band at 340 nm and a weak positive band at 440 nm which could indicate coralyne-mediated disproportionation of an RNA (Poly rU)•(Poly rA) duplex. Coralyne concentration was 0.50 molar equivalents per base pair.

Poly (rA) single strand in the presence of coralyne appears to associate into a self structure. Recently published work by Chaires and coworkers have shown that coralyne binds strongly to a Poly (rA) sequence with an association constant of 1.8 X 10^6 M^{-1} [116]. In addition, Poly (rA) associates into a secondary structure in the presence of coralyne (T_m = 60ºC). Our work shows that a sample of Poly (rA) in the absence of coralyne shows no CD bands above 270 nm. However, upon addition of coralyne, significant changes in the CD spectrum occur at 220 nm, 249 nm and 270 nm. In addition, self-structure formation of a single strand Poly (rA) is supported by the appearance of positive bands at 340 nm and 440 nm which suggest coralyne binding to the single stranded RNA structure (Figure 2.15).

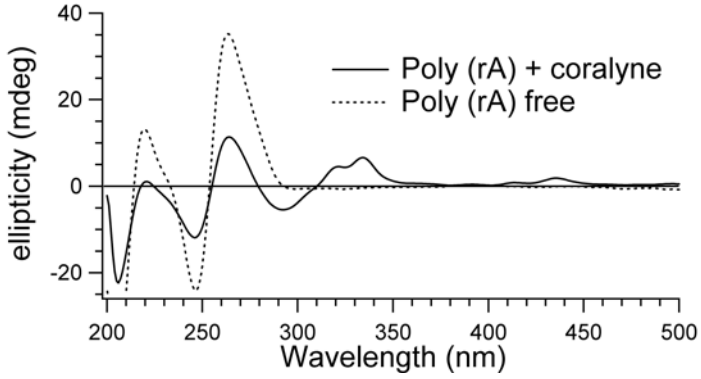


Figure 2.15. CD spectra showing Poly (rA) interactions with coralyne.Circular dichroism spectra as a function of wavelength show that a Poly (rA) sample upon addition of coralyne (solid) exhibits strong positive bands at ~ 340 nm and ~ 440 nm. Dashed line is Poly (rA) sample in the absence of coralyne. Coralyne concentration was 0.50 molar equivalents per base.

According to our results, coralyne binding to an RNA triplex is less favorable than its binding to DNA structures under the condition of our studies. Competition dialysis experiments also support this conclusion [116]. RNA duplex disproportionation is not promoted by coralyne, as cooling samples of (Poly rU)•(Poly rA) after heat cycling

does not lead to the appearance of bands associated with triplex intercalation. Further work is warranted in the area of coralyne interactions with RNA structures.

2.4. CONCLUDING REMARKS

We have demonstrated that nucleic acid binding by a small molecule, such as coralyne, is a powerful means to control DNA secondary structure. However, we have also shown that the kinetics and thermodynamics of DNA structure formation in the presence of small molecule intercalators can be complex. Even our rather minimalist system composed of $(dA)_n \cdot (dT)_n$ duplexes has revealed several aspects of DNA secondary structure formation in the presence of coralyne that depend upon strand length. Part of this complexity may result from the fact that coralyne interacts with $(dA)_n \cdot (dT)_n$ duplexes in a manner that is different from its interaction with $(dA)_n \cdot (dT)_n \cdot (dT)_n$ triplexes, which may also be distinct from its interaction with the $(dA)_n$ self-structure. If we consider that the interaction between coralyne and each of these DNA secondary structures is most likely defined by different enthalpies and entropies of binding, it is then perhaps not as surprising that coralyne binding can produce a complex relationship between temperature and DNA secondary structure. One aspect of DNA-coralyne interactions that we have not explored in the present work is the possible effect of coralyne concentration on DNA secondary structure equilibrium. It is altogether possible that altering intercalator concentration could be a means to control secondary structure equilibrium at a particular temperature.

CHAPTER 3

ENZYMATIC BEHAVIOR BY INTERCALATING MOLECULES IN A TEMPLATE DIRECTED LIGATION REACTION

3.1. INTRODUCTION

Since the discovery of catalytic RNA two decades ago [13, 126], much attention has focused on the hypothesis that an early form of life used nucleic acids for both information storage and catalysis, before the advent of proteins [12, 127]. However, it is still a mystery how the first nucleic acid polymers assembled and replicated, as these tasks are carried out by protein enzymes in contemporary life. Decades of research have lead to the inescapable conclusion that Watson-Crick base pairing alone does not sufficiently stabilize the assembly of mononucleotides on a template strand in aqueous solution to allow spontaneous self-replication [14, 52]. Investigations of non-natural mononucleotide coupling chemistries and chemical activation have proven more successful than attempts to condense the natural mononucleotide triphosphates on single stranded DNA or RNA templates [20, 40, 42, 47, 128, 129]. Nevertheless, a prebiotically-plausible method to bridge the gap from small molecules to self-replicating RNA-like polymers has not been found.

A multitude of small molecules must have existed on prebiotic Earth, and only a small subset of these molecules is likely represented by the building blocks of biopolymers in contemporary life (e.g. sugars, amino acids, nucleotide bases). In 2000, Hud and Anet proposed a hypothesis which presents a new approach to the origin of nucleic acid polymers. The *molecular midwife* hypothesis proposes that small molecules

that intercalate the bases of nucleic acids could have also been present on the prebiotic Earth and acted as nanoscale surfaces upon which nucleotide bases (RNA or RNA-like) stacked to form ordered assemblies [57]. Thus, the molecular midwife hypothesis represents a potential bridge between a prebiotic small-molecule world and the RNA World.

In this work here, we have shown that a small molecule that intercalates the bases of DNA and RNA can increase the template-directed coupling rate of short oligonucleotides by three orders of magnitude [85]. Several of these molecules work together in a cooperative manner to function, in essence, as a concentration-dependent multimolecular "enzyme". These results support the recently made hypothesis that an intercalating molecule could have acted as a "molecular midwife" that facilitated the replication of information-containing polymers before the existence of the RNA World, as well as the replication of RNA itself, at least in the early stages of the RNA World [57].

3.2. EXPERIMENTAL PROCEDURES

3.2.1. Sample Preparation

Substrate oligodeoxynucleotides were synthesized on an automated synthesizer using the phosphoramidite coupling chemistry. Synthesis of 3'-phosphorothioate-$(dT)_3$ was accomplished using a 3'-phosphate controlled pore glass support (Glen Research) where the oxidation reagent normally added during the first nucleotide coupling cycle was replaced by a sulfurizing reagent (Glen Research). The 5'-iodo-$(dT)_4$ substrate oligo was synthesized using a 5'-iodothymidine phosphoramidite reagent (Glen Research). Following deprotection, substrate oligonucleotides were purified by reverse phase HPLC on a C_{18}, semi-preparative column and characterized using mass spectrometry analysis. Template strand oligonucleotides were purified by passage over a 1 m G-15 column (Pharmacia). Stock solutions of oligonucleotides were prepared by resuspending freeze-dried purified samples in deionized H_2O. Oligonucleotide concentrations were determined spectrophotometrically.

3.2.2. Radiolabeling Assay

The 3'-phosphorothioate-$(dT)_3$ substrate was radioactively labelled with ^{32}P-phosphate at the 5'-end by diluting 3'-phosphorothioate-$(dT)_3$ from a stock solution to 50 μM strand in 100 μl of T4 polynucleotide kinase buffer (New England Biolabs). 30 units of T4 polynucleotide kinase (New England Biolabs) were added to the buffered DNA solution. 3 μl γ-^{32}P-ATP (ICN), 100 μCi/μl, was then added to the solution and allowed to incubate at 37°C for 30 min.

3.2.3. Proflavine Stock Preparation

Proflavine (hemisulfate salt) was purchased from Sigma. Stock solutions of proflavine were prepared by dissolving the solid proflavine salt in deionized H_2O. Stock solution concentrations were determined spectrophotometrically using the extinction coefficient: $\varepsilon_{445} = 38{,}900\ M^{-1}\ cm^{-1}$.

3.2.4. Template Directed Ligation Reactions

Reactions were carried out in 100 μl volumes in a solution containing 10 mM Tris buffer (pH 8.2); 10 mM NaCl; 100 mM 2-mercaptoethanol. The substrate 5'-iodo-$(dT)_4$, the substrate ^{32}P-labeled-3'-phosphorothioate-$(dT)_3$, and the template $(dA)_{16}$ were each added to the reaction buffer to a final concentration of 1.0 μM in strand. The presence of 2-mercaptoethanol in the reaction buffer was necessary to prevent dimerization of the 3'-phosphorothioate-$(dT)_3$ substrate. Ligation reactions were stopped by plunging reaction test tubes into liquid nitrogen and freeze-drying.

3.2.5. Product analysis

Freeze-dried reaction samples were resuspended in 10 μl of a 8 M urea solution and loaded on to a denaturing, 30% polyacrylamide gel (19:1::acrylamide:bisacrylamide). Gels were subject to electrophoresis at a constant power of 70 W for 5 hours. The relative yield of $(dT)_7$ product for each reaction was determined by imaging the gel on a Fuji Phosphor Imager and quantifying the integrated intensity of gel bands in each lane that corresponded to the $(dT)_7$ ligation product using the software package Image Gauge V3.12. Background correction was accomplished by subtracting from all reaction

samples the integrated intensity of an area in a control lane run with only ^{32}P-end-labeled $(dT)_3$.

3.3. RESULTS AND DISCUSSION

3.3.1. Proflavine Increases Ligation Rate of Oligonucleotides by Three Orders of Magnitude

We have conducted a series of experiments to test whether intercalation in present day nucleic acids can facilitate nucleic acid template-directed synthesis. Our experimental test system involves suitably modified forms of the short oligonucleotides, $(dT)_3$ and $(dT)_4$ (Figure 3.1). The chemistry used to couple these oligonucleotides makes use of an iodine atom as a leaving group on 5'-iodo-$(dT)_4$ and leads to covalent bond formation with the sulfur atom of 3'-phosphorothioate-$(dT)_3$ [130, 131].

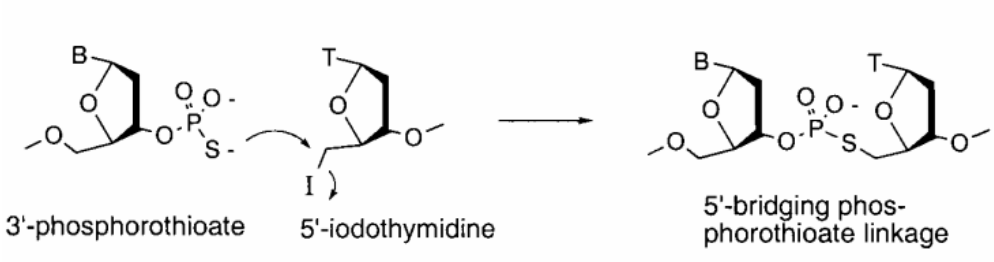


Figure 3.1. Phosphorothioate-iodo coupling chemistry. An oligonucleotide with a 3'-phosphorothioate group attacks an oligonucleotide with a 5'-iodo group leading to the subsequent formation of covalent phosphorothioate linkage (Adapted from reference 14).

A graphical representation of this ligation test system is shown in Figure 3.2A. The intercalator used is a planar tricyclic cationic molecule commonly known as proflavine (Figure 3.2B, 3.2C), which closely matches the shape of a Watson-Crick base pair (Figure 3.2D, 3.2E).

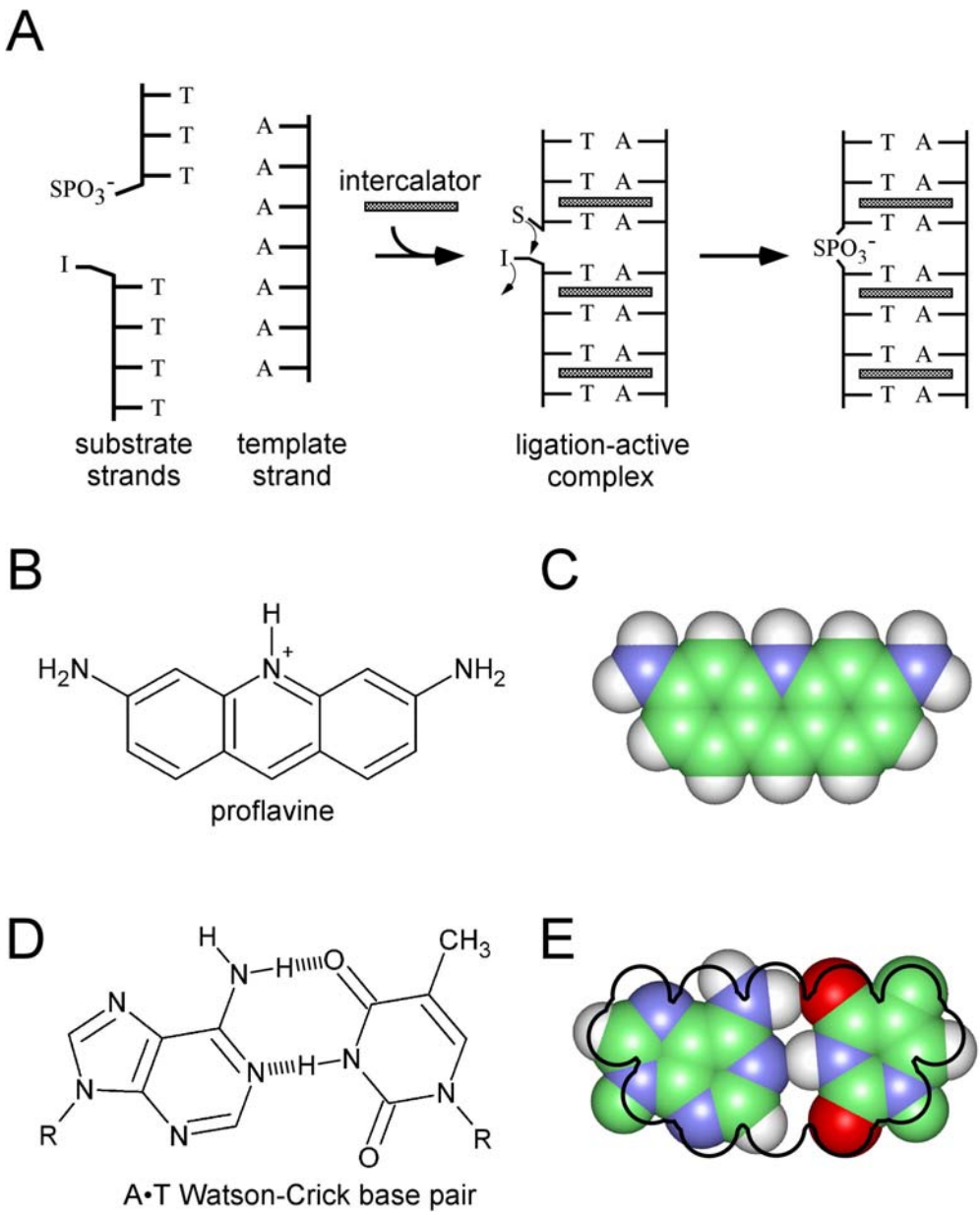


Figure 3.2. A schematic representation of the ligation test system with applicable molecular structures. A) A template strand in solution with substrate strands in which the substrate strands are sufficiently short such that the equilibrium amount of substrate strands bound to the template strand is extremely small. The addition of an intercalating molecule to the solution facilitates the formation of a duplex between the template strand and substrate strands with Watson-Crick complementary sequences. Chemical ligation is used to join the backbones of substrate strands aligned along the template strand. B and C) The chemical structure and space-filling model of proflavine. D and E) The chemical structure and space-filling model of the Watson-Crick A•T base pair. A black outline of the proflavine van der Waals surface is superimposed on the space-filling model of the A•T base pair to illustrate the close match between the shapes of these molecular structures.

By labeling the 3'-phosphorothioate-$(dT)_3$ substrate on its 5' end with ^{32}P-phosphate, we are able to follow product formation as a function of intercalator concentration by quantification of the $(dT)_7$ product after polyacrylamide gel electrophoresis. An image of a gel for a set of ligation experiments with increasing concentrations of proflavine is presented in Figure 3.3.

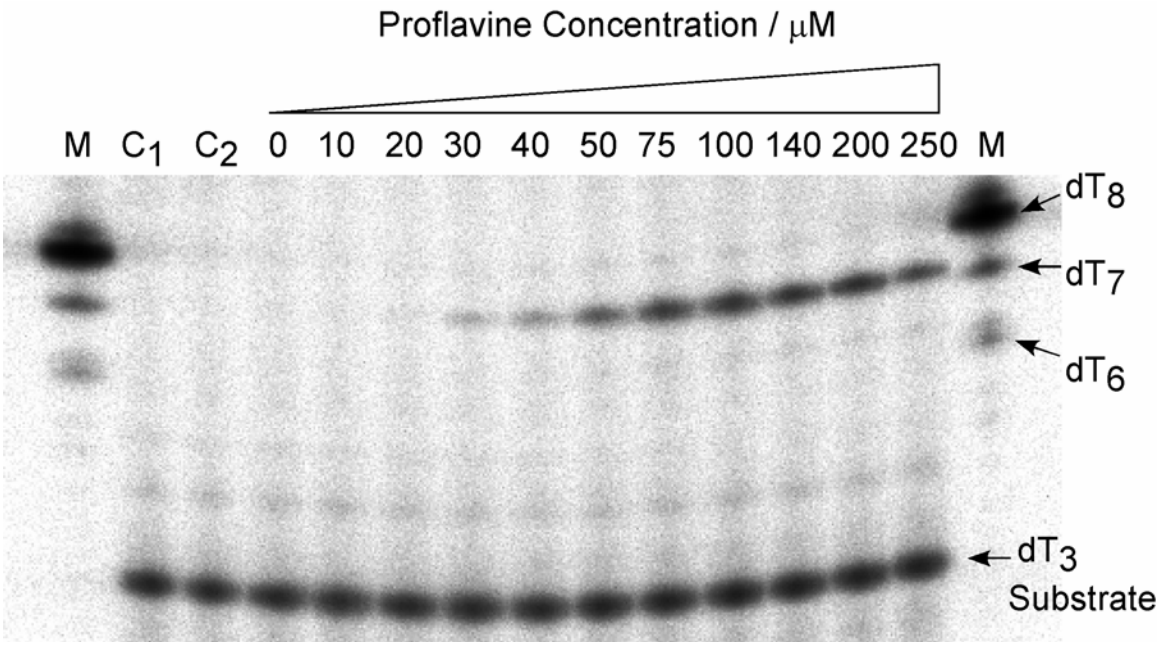


Figure 3.3. Gel electrophoresis of the ligation system. Denaturing polyacrylamide gel electrophoresis analysis that illustrates the effect of proflavine on the ligation yield of 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ using $(dA)_{16}$ as a template strand. Lane C1: Only ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$. Lane C2: Substrates ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$. Lanes labelled 0 to 250: ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$, 5'-iodo-$(dT)_4$, template strand $(dA)_{16}$, and proflavine at a concentration corresponding to the number above the lane, in units of μM. All reaction mixtures were incubated for 24 h at 4ºC. Lane M: Molecular weight marker bands of $(dT)_8$, $(dT)_7$ and $(dT)_6$.

This gel shows that the $(dT)_7$ product is virtually undetectable for reactions containing the substrates $(dT)_3$, $(dT)_4$ and the $(dA)_{16}$ template strand. The addition of proflavine to the reaction mixture of $(dT)_3$ and $(dT)_4$ produces a detectable increase in the $(dT)_7$ ligation product yield, even in the complete absence of the template. The significance of this result will be discussed later. A far more dramatic increase in $(dT)_7$ yield occurs when both the $(dA)_{16}$ template and proflavine are present (Figure 3.3). Quantification of gel band intensities (Table 3.1) shows that proflavine catalyzes the ligation rate of $(dT)_3$ and $(dT)_4$ by three orders of magnitude over reactions relying on only the $(dA)_{16}$ template strand to organize the substrates. These results are consistent with proflavine promoting the formation of a $(dT)_3$, $(dT)_4 \cdot (dA)_{16}$ duplex that acts as a ligation complex where the reactive ends of the $(dT)_3$ and $(dT)_4$ substrates can meet. The ligation product is a phosphorothioate-linked analog of $(dT)_7$ (Figure 3.2A). The importance of Watson-Crick base pairing in the proflavine-catalyzed ligation reaction is illustrated by the fact that product yield drops significantly when DNA templates with sequences other than $(dA)_{16}$ are used with the $(dT)_3$ and $(dT)_4$ substrates (Table 3.1).

3.3.2. Proflavine Molecules Display Cooperativity in Binding to Ligation Complex

A plot of the $(dT)_7$ ligation rate on a $(dA)_{16}$ template demonstrates that the proflavine-catalyzed reaction is cooperative in proflavine concentration to approximately 100 μM proflavine (Figure 3.4). A least-squares-fit of this data by the Hill equation indicates that at least three proflavine molecules bind cooperatively to the substrate and template strands, each with a binding constant of around 60 μM, to create the active ligation complex.

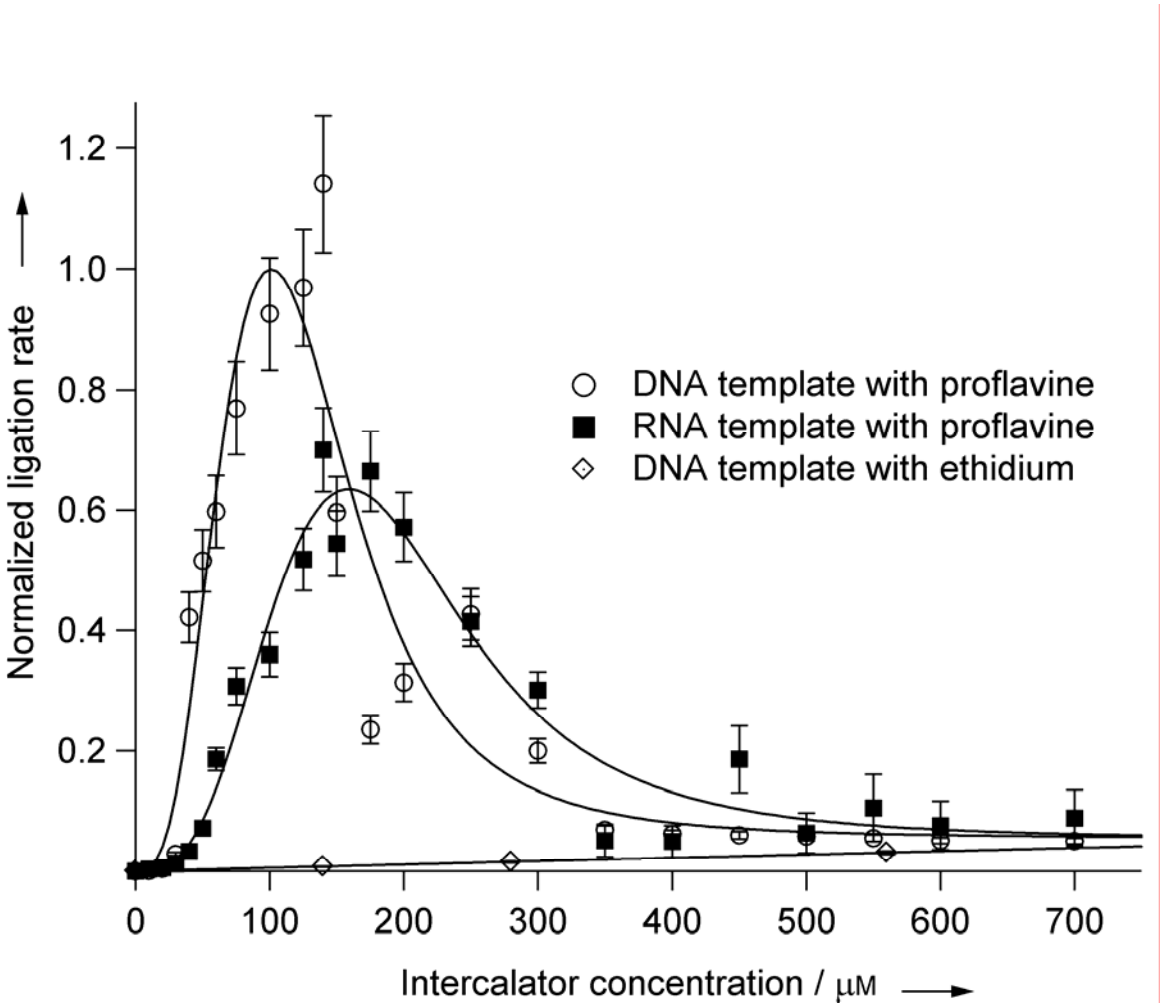


Figure 3.4. Plots of the relative ligation rates (R) for formation of the $(dT)_7$ product as a function of template strand, $(dA)_{16}$ or $(rA)_{16}$, intercalator species, proflavine or ethidium, and intercalator concentration. Rates shown have been normalized to the maximum of the fit of the data for proflavine with the DNA template. Substrates, 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$, and template strands were 1.0 μM in oligonucleotide strand for all reactions. The reaction mixtures were incubated for 24 h at 4ºC. The error bars show known sources of error only. A few data points, for unknown reasons, show unexpectedly large deviations from the fitted curves; omitting these points does not change the fits appreciably.

Table 3.1: Quantitative analysis of ligation test system.

Template[a]	Intercalator[b]	Ligation rate[c]	Half Max[d]
-	-	<15	NA
$(dA)_{16}$	-	<15	NA
$(dA)_{16}$	Proflavine	10,000 ± 1,000	51 μM
$(dAATA)_4$	Proflavine	5,100 ± 600	ND
$(dN)_{16}$	Proflavine	~70	ND
$(dA)_{16}$ (25°C)	Proflavine	6,300 ± 700	ND
$(rA)_{16}$	Proflavine	7,100 ± 800	87 μM
$(dA)_{16}$	Ethidium Bromide	~35	>500 μM
-	Proflavine	~70	ND

[a] Template $(dAATA)_4$ is d(AATAAATAAATAAATA); template $(dN)_{16}$ is d(GATCCGAATTCACGTG). [b] Intercalator concentrations were 140 μM, where an intercalator is listed. [c] Ligation rates were determined based upon radioactive decay counts from gel bands and have been normalized with respect to the highest ligation rate experiment, which has been scaled to 10,000. [d] Half Max, concentration of intercalator at which product yield is one half of the maximum ligation rate; NA, not applicable; ND, not determined. All reactions contained both the $(dT)_3$ and $(dT)_4$ substrates as described in text. All experiments were carried out at 4°C, unless otherwise indicated. Ligation reaction time was 24 h for all experiments.

According to the nearest neighbor exclusion principle, the bases of nucleic acid duplexes can only bind one intercalating molecule per two base pairs [67, 132]. Thus, the substrates $(dT)_3$ and $(dT)_4$, when forming a duplex with a $(dA)_n$ template strand, would be expected to bind one and two proflavine molecules respectively (Figure 3.2A), for a total of three molecules, in agreement with our experimental data.

The 1000-fold increase in the rate of $(dT)_7$ ligation product formation in a solution containing 140 μM proflavine implies that proflavine reduces the overall free energy barrier for ligation by approximately 3.8 kcal/mol at 277 K. When the same reaction was carried out at 298 K, the rate of product formation was reduced by a factor of 0.63 with respect to the rate at 277 K (Table 3.1). Kool and co-workers have shown that, for a stable nucleic acid assembly, the phosphorothioate ligation reaction rate increases with temperature over this range [131]. Thus, the reduction in product formation rate observed in our system with increased temperature must be due to a reduction in the concentration of the duplex structure containing three intercalated proflavine molecules, consistent with the expected negative entropy for the formation of such a complex. However, the analysis is complicated by the presence of a second intercalation complex evident at still higher proflavine concentration.

3.3.3. Decline in Ligation Activity is Observed at Higher Proflavine Concentrations

Under the conditions used, the maximum rate of proflavine-catalyzed $(dT)_7$ ligation on the DNA template $(dA)_{16}$ is achieved at around 100 μM proflavine, but it does not remain at a plateau. Instead, it decreases smoothly after the point of maximum yield to reach a much lower constant rate at approximately 600 μM proflavine (Figure 3.4).

The shape of the $(dT)_7$ ligation product curve for proflavine concentrations from 0 to 700 μM proflavine is consistent with an additional four more proflavine molecules binding to the reaction complex with a weaker binding constant (~160 μM) than the first three proflavine molecules that assemble the catalytically-active complex. The decrease in ligation rate for high proflavine concentrations indicates that the higher-bound proflavine assembly is a much less catalytically active complex than the three-proflavine complex. It is possible that the secondary set of proflavine binding sites arrange the $(dT)_3$ and $(dT)_4$ oligonucleotides such that their reactive groups are too far away from each other for bond formation, or that high proflavine concentrations induce the $(dA)_{16}$ template to dimerize [8]. In any case, the significant decline in reaction rate upon the binding of more than three proflavine molecules fits a cooperative phenomenon.

In Figure 3.4 we also present results from proflavine-catalyzed ligation of $(dT)_3$ and $(dT)_4$ on the RNA template $(rA)_{16}$. The overall results are similar to those with the analogous DNA template, except that the curve is shifted to higher proflavine concentration, indicating that the intercalation complex is somewhat less favorable than with the DNA template. This result shows the interplay that exists between a small molecule intercalator and backbone structure, even though an intercalator such as proflavine is expected to have minimal direct contacts with the backbone (Figure 3.2E).

Ethidium bromide, a common fluorescent intercalator, was also investigated in our ligation test system. Far less $(dT)_7$ ligation product was observed in comparison to the same reaction with proflavine (Figure 3.4). The binding constants of proflavine and ethidium for a DNA duplex are very similar [20, 69, 91]. Thus, the ability for an intercalating molecule to act as a midwife must also depend on the shape of the molecule,

and not simply its binding constant. Proflavine has three linearly fused aromatic rings, whereas ethidium has its three aromatic rings angularly fused, and it also has a pendant phenyl group not present in proflavine. This hydrophobic phenyl group would tend to increase the binding of ethidium in an aqueous medium, but might well be detrimental to the ligation reaction itself.

As noted above, a small but distinct increase in the rate of $(dT)_7$ ligation product formation over background was detected in a reaction mixture containing 140 μM proflavine but no $(dA)_{16}$ template strand (Table 3.1). A small increase in the template-free ligation rate of $(dT)_3$ and $(dT)_4$ by proflavine is of interest because it shows that an intercalator, perhaps through non-specific stacking interactions with the terminal bases of $(dT)_3$ and $(dT)_4$, can create a small equilibrium amount of a ligation-active complex. This means that DNA- and RNA-like polymers could have been synthesized *de novo* by intercalators at low rates without the requirement for pre-existing templates. Once this occurs, the system could become autocatalytic if complementary Watson-Crick bases were both present as activated monomers, since the spontaneous emergence of template strands would greatly enhance the production of complementary strands in the presence of the proper intercalator.

The plots of $(dT)_7$ ligation rates shown in Figure 3.4 can also be viewed as plots of the rates of enzyme-catalyzed reactions as functions of enzyme concentrations. The rate of an enzyme-catalyzed reaction typically increases linearly with enzyme concentration (i.e. first order in enzyme concentration). In contrast, the cooperative increase in $(dT)_7$ ligation product with proflavine concentration indicates that the three proflavine molecules of the active complex are working together. Thus, the small molecule

proflavine can be viewed as a cooperative-concentration-dependent multimolecular enzyme. This fact has significant implications regarding the possible utility of small planar molecules and the role of intercalation in the early stages of life [57].

3.3.4. HPLC Purification and MS Analysis of the Modified Substrates

3.3.4.1. 3'-phosphorothioate-(dT)$_3$

Figure 3.5 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time for the purification of 3'-phosphorothioate-(dT)$_3$ oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 5% acetonitrile, 50 min = 25% acetonitrile. Flow rate was 4.6 ml/min.

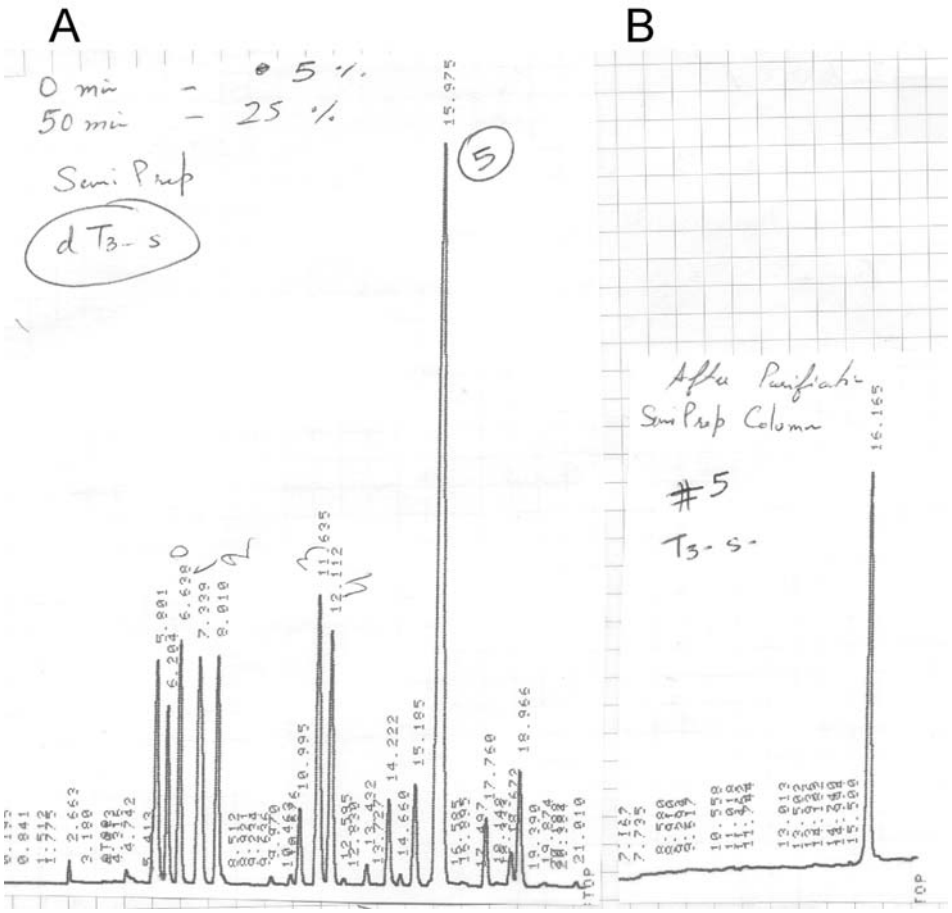


Figure 3.5. HPLC purification of 3'-phosphorothioate-(dT)$_3$. (A) Chromatograph of a crude sample where fractions (1-5) were collected. (B) Fraction # 5 run on the column a second time to confirm purity.

Mass spectrum of a purified 3'-phosphorothioate-$(dT)_3$ sample is shown in Figure 3.6. Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 946.6 g/mol. M-H peak is observed at 945.2 g/mol. We also observe a peak at 472.1 g/mol which is the M-2H charged species. The significant peak at a M/Z (mass to charge) ratio of 849.2 g/mol is possibly the oligonucleotide with a 3'-OH instead of a 3'-phosphorothioate. This analysis cannot be used as a quantitative comparison because different compounds have different ionizing potentials.

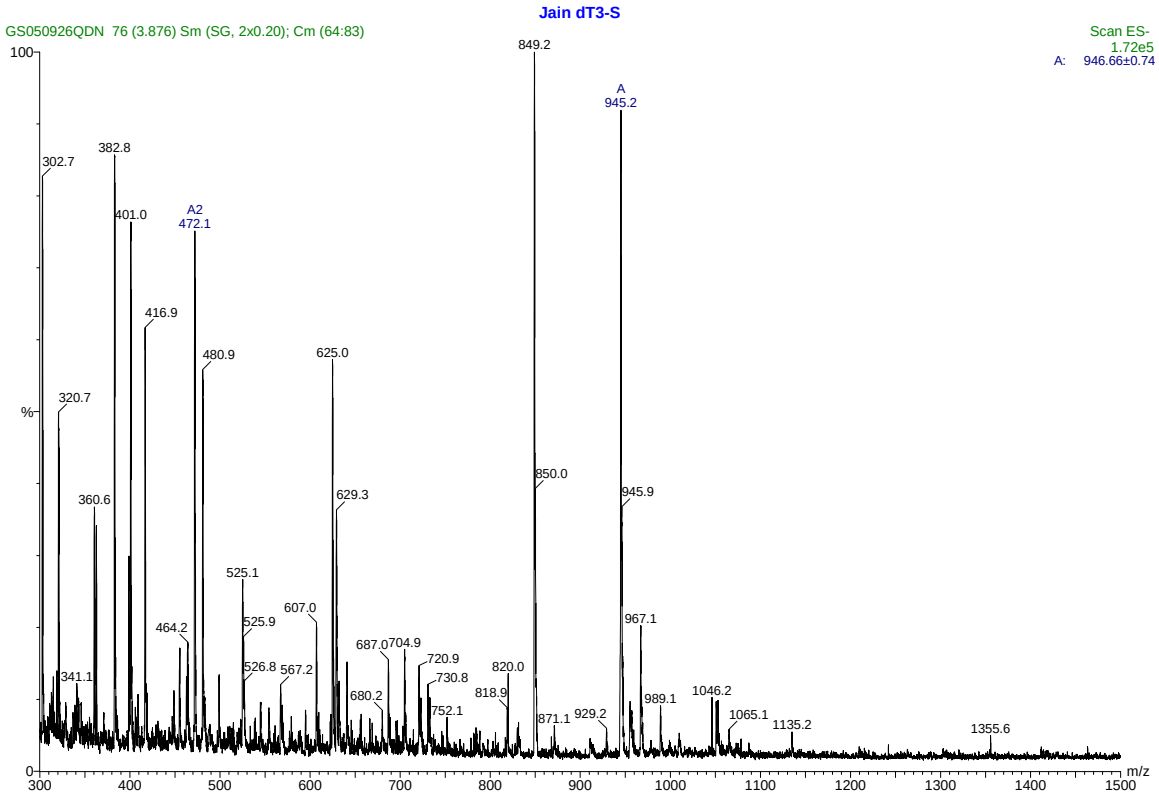


Figure 3.6. Mass Spectrum of a 3'-phosphorothioate-$(dT)_3$ oligonucleotide.

3.3.4.2. 5'-iodo-(dT)$_4$

Figure 3.7 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time for the purification of 5'-iodo-(dT)$_4$ substrate. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 8% acetonitrile, 50 min = 25% acetonitrile. Flow rate was 4.6 ml/min.

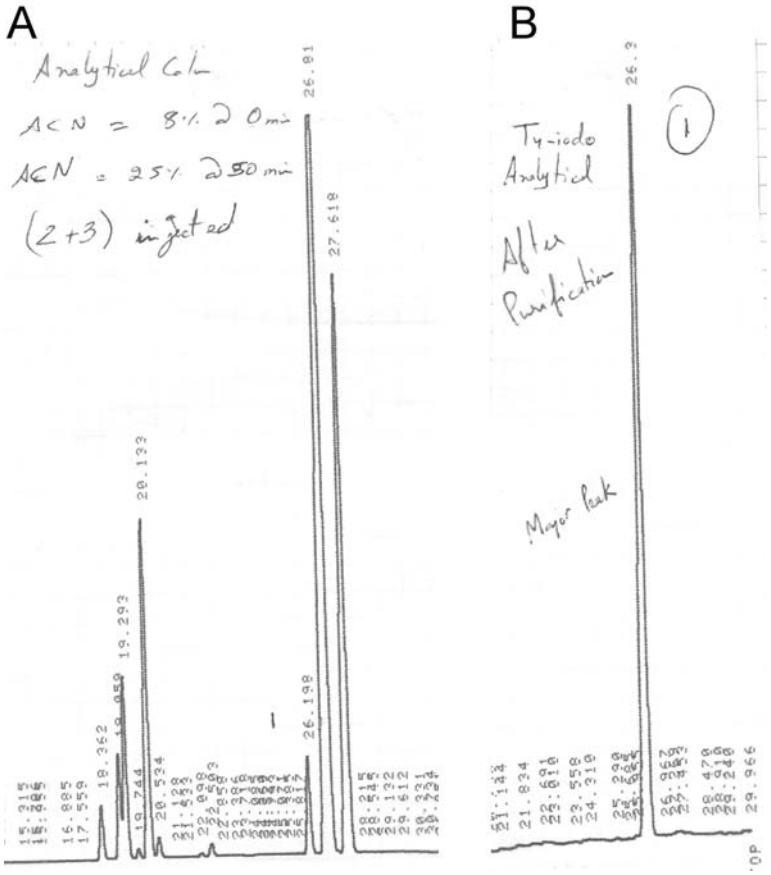


Figure 3.7. HPLC purification of 5'-iodo-(dT)$_4$. (A) Chromatograph of a crude sample where fraction # 1 (26.8 min elution time) was collected. (B) Fraction # 1 was run on the column a second time to confirm purity.

Mass spectrum of a purified 5'-iodo-$(dT)_4$ sample is shown in Figure 3.8. Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 1264.7 g/mol. M-H peak, labelled as B in Figure 3.8, is observed at 1263.2 g/mol. We also observe a peak, labelled B2 at 631.2 g/mol which is the M-2H charged species. Peak A at an M/Z (mass to charge) ratio of 1135.2 g/mol is potentially the oligonucleotide with a 5'-OH instead of a 5'-iodo group.

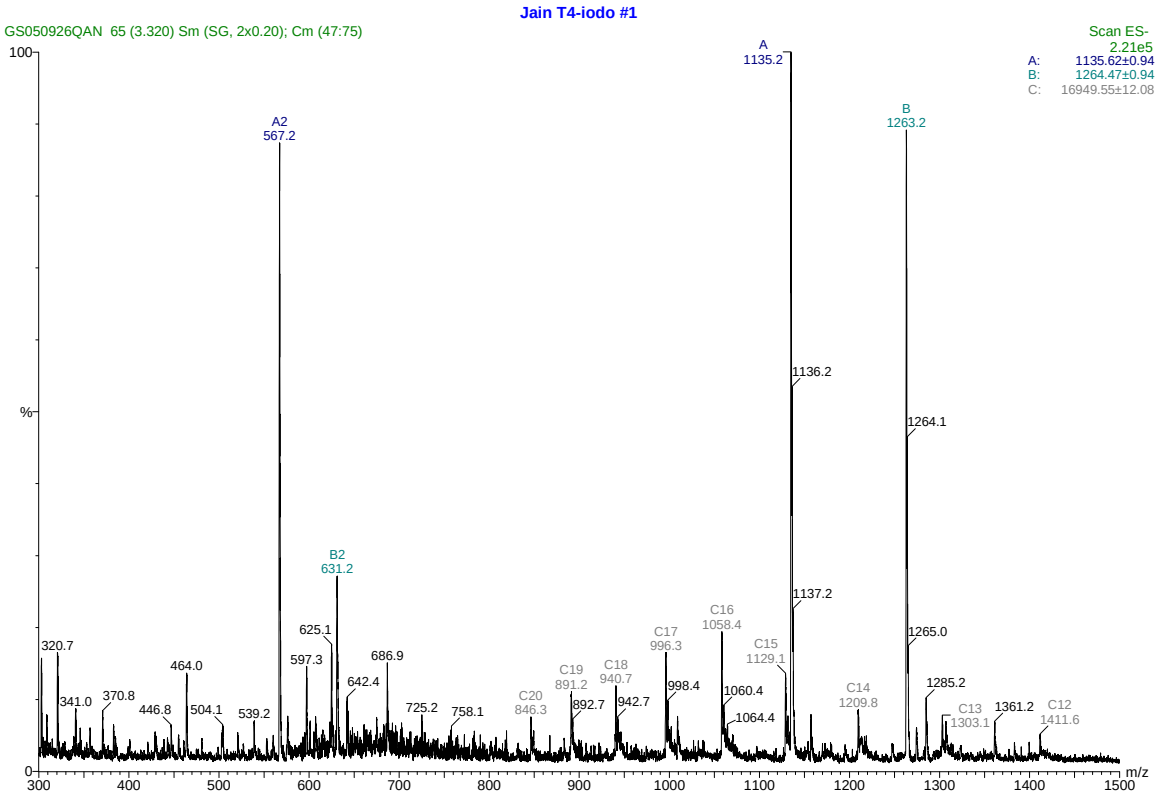


Figure 3.8. Mass Spectrum of a 5'-iodo-$(dT)_4$ oligonucleotide.

3.4. CONCLUDING REMARKS

In conclusion, our demonstration that an intercalating molecule can greatly increase the efficiency of a template-directed ligation reaction has important implications for contemporary nucleic acid chemistry, as well as potential implications concerning the mechanism of nucleic acid synthesis in early life. For over thirty years researchers have sought to improve the yield of protein-free template-directed nucleic acid ligation reactions. Past efforts have included careful sequence design; exhaustive exploration of solution conditions; use of templates with non-natural backbones; and the development of novel substrate linkage chemistries [3]. Our results demonstrate that the simple act of adding an intercalating molecule to a ligation reaction can have a huge effect on improving coupling efficiency. In regards to early life, there has also been much speculation concerning the possible role of inorganic surfaces in the origin of life [41, 133-135], as the collection of materials on surfaces could serve as a means to concentrate and spatially organize the molecular components necessary for life. However, as we have illustrated here, a relatively simple molecule with a flat surface could have accomplished these tasks in a much more versatile way than a solid macroscopic surface. Molecules that intercalate DNA and RNA duplexes do so in part because their shapes match those of Watson-Crick base pairs. In the same way, molecules that could have acted as molecular midwives in the assembly and replication of the first informational polymers may have played a significant role in selecting the nucleotide bases due to their ability to form structures that matched the structure of the midwife's surface.

CHAPTER 4

EXPLORING THE EFFECT OF SEVERAL PARAMETERS ON SMALL MOLECULE MEDIATED ASSEMBLY OF NUCLEIC ACIDS

4.1. INTRODUCTION

Small molecule binding to nucleic acids represents a new and powerful method for enhancing the ligation rate of short oligonucleotides in a template-directed reaction. We recently provided initial experimental evidence that small molecule intercalation greatly enhances the coupling of oligonucleotide substrates in a template-directed reaction [85]. In the present study we have investigated the effect of altering several parameters in midwife-mediated ligation reactions. We demonstrate that the catalytic activity of midwife molecules is strongly dependent upon solution conditions, template length, sequence mismatches and midwife chemical structure. Crowding agents, such as polyethylene glycol (PEG), can also lead to appreciable changes in the ligation activity. The results presented here also illustrate that the molecular midwife approach can be extended beyond Watson-Crick base pairing to artificial replicating systems that use non-standard base pairs. We have investigated the effect of cations in solution on proflavine-mediated ligation reactions. We have also explored whether proflavine can assemble a hybrid ligation complex with DNA substrates and varying types of RNA templates. Basic guidelines based upon these studies should facilitate future developments that more fully exploit the molecular midwife approach to template-directed ligation and replication of nucleic acids.

4.2. EXPERIMENTAL PROCEDURES

4.2.1. Synthesis of Oligonucleotides

Substrate oligodeoxynucleotides were synthesized on an automated synthesizer using the phosphoramidite coupling chemistry. Synthesis of 3'-phosphorothioate-$(dT)_3$ was accomplished using a 3'-phosphate controlled pore glass support (Glen Research) where the oxidation reagent normally added during the first nucleotide coupling cycle was replaced by a sulfurizing reagent (Glen Research). The 5'-iodo-$(dT)_4$ substrate oligo was synthesized using the commercially available 5'-iodothymidine phosphoramidite reagent (Glen Research). Following deprotection, substrate oligonucleotides were purified by HPLC. All DNA template strand oligonucleotides were purchased from IDT and desalted over a standard desalting column. Template strand $(rA)_7$ and 2'-O-Me $(rA)_7$ were purchased from Dharmacon and deprotected according to the manufacturer's protocols. Stock solutions of oligonucleotides were prepared by resuspending freeze-dried purified samples in deionized H_2O. Oligonucleotide concentrations were determined spectrophotometrically.

4.2.2. Radiolabeling Substrate 3'-phosphorothioate-$(dT)_3$

The 3'-phosphorothioate-$(dT)_3$ substrate was radioactively labelled with ^{32}P-phosphate at the 5'-end by diluting 3'-phosphorothioate-$(dT)_3$ from a stock solution to 50 μM per strand in 100 μl of T4 polynucleotide kinase buffer (New England Biolabs). 30 units of T4 polynucleotide kinase (New England Biolabs) were added to the buffered DNA solution. 3 μl γ-^{32}P-ATP (ICN), 100 μCi/μl, was then added to the solution and allowed to incubate at 37°C for 30 min.

4.2.3. Ligand Stock Solutions

Proflavine hemisulfate, coralyne chloride, actinomycin D, distamycin A•HCl, BePI, daunomycin and DAPI were purchased from Sigma. Ethidium bromide and acridine orange were obtained from Fisher and Molecular Probes, respectively. Stock solutions were prepared by dissolving the solid salts in deionized H_2O, except for BePI (DMSO), acridine orange and actinomycin (methanol). Stock concentrations were determined spectrophotometrically using the following extinction coefficients: ε_{444} = 38,900 M^{-1} cm^{-1} (proflavine), ε_{420} = 14,500 M^{-1} cm^{-1} (coralyne), ε_{482} = 5,500 M^{-1} cm^{-1} (ethidium), ε_{303} = 37,000 M^{-1} cm^{-1} (distamycin) and ε_{489} = 67,000 M^{-1} cm^{-1} (acridine orange), ε_{344} = 27,000 M^{-1} cm^{-1} (DAPI), ε_{361} = 5,500 M^{-1} cm^{-1} (BePI), ε_{480} = 11,500 M^{-1} cm^{-1} (daunomycin), ε_{441} = 26,400 M^{-1} cm^{-1} (actinomycin).

4.2.4. Template-Directed Ligation Assay

Reactions were carried out in a solution containing 10 mM Tris buffer (pH 8.2), 10 mM NaCl, and 100 mM 2-mercaptoethanol. The substrate 5'-iodo-$(dT)_4$, the substrate ^{32}P-labeled-3'-phosphorothioate-$(dT)_3$, and the template $(dA)_n$ were each added to the reaction buffer to a final concentration of 1.0 μM in strand. The presence of 2-mercaptoethanol in the reaction buffer was necessary to prevent dimerization of the 3'-phosphorothioate $(dT)_3$ substrate. All reaction mixtures were incubated at 4ºC for 24h. Ligation reactions were stopped by plunging reaction test tubes into liquid nitrogen and freeze-drying.

4.2.5. Product Analysis and Quantification

Freeze-dried reaction samples were resuspended in 10 μl of a 7 M urea solution and loaded on to a denaturing, 30% polyacrylamide gel (19:1::acrylamide:bisacrylamide). Gels were subject to electrophoresis at a constant power of 70 W for 5 hr. The relative yield of $(dT)_7$ product for each reaction was determined by imaging the gel on a Fuji Phosphor Imager and quantifying the integrated intensity of gel bands in each lane that corresponded to the $(dT)_7$ ligation product using the software package Image Gauge V3.12. Background correction was accomplished by subtracting, from all reaction samples, the integrated intensity of an area in a control lane run with only ^{32}P-end-labelled $(dT)_3$.

4.2.6. Circular Dichroism Spectroscopy

Circular dichroism experiments were performed at 4ºC using a JASCO J-810 CD spectropolarimeter equipped with a Peltier temperature control device. Spectra were acquired using 10 mm and 5 mm path length cells (Helma). Sample shown in Figure 4.8B contained 6 μM per strand $(dA)_7$:$(dT)_7$ duplex, 10 mM Tris buffer (pH 8.2) and 60 μM proflavine. Sample shown in Figure 4.8C contained 120 μM per base pair $(dA)_7$:$(dT)_7$ duplex, 10 mM Tris buffer (pH 8.2) and 60 μM proflavine. Ionic strength was increased by titrating NaCl and $MgCl_2$ stock solutions into the proflavine:DNA sample.

4.3. RESULTS AND DISCUSSION

4.3.1. Effect of Template $(dA)_n$ Length on Ligation Activity

We have previously shown that with the oligonucleotide $(dA)_{16}$ as a template, the proflavine-catalyzed ligation rate of the $(dT)_3$ and $(dT)_4$ substrates is enhanced by more than three orders of magnitude [85]. However, complete matching of the substrates to the template strand without any substrate sliding or overhanging effects occurs only in the presence of a $(dA)_7$ template. In order to explore the effects of template length on ligation assay, four template lengths, $(dA)_5$, $(dA)_7$, $(dA)_{12}$, and $(dA)_{16}$, were compared in this study, with the substrates being the same $(dT)_3$ and $(dT)_4$ oligonucleotides described above. The template concentrations were kept constant at 7 μM per base for all four templates. Product ligation rates for the different template lengths were measured for selected proflavine concentrations from 0 to 1000 μM (Figure 4.1).

Ligation-rate profiles measured for $(dA)_5$, $(dA)_7$, $(dA)_{12}$, and $(dA)_{16}$ templates all exhibit an initial increase in yield that is cooperative in proflavine concentration (Figure 4.1). Additionally, reaction rates for the three longest templates exhibit a smooth decline after reaching a maximum at approximately 125 μM proflavine. There are, however, definite differences between the ligation rate profiles for templates $(dA)_7$, $(dA)_{12}$, and $(dA)_{16}$. The maximum ligation rates measured as a function of proflavine concentration followed the trend $(dA)_7 > (dA)_{12} > (dA)_{16}$. The decline in the ligation rate of reactions containing template $(dA)_7$ at higher proflavine concentrations was also somewhat more gradual than that of reactions containing templates $(dA)_{12}$ and $(dA)_{16}$.

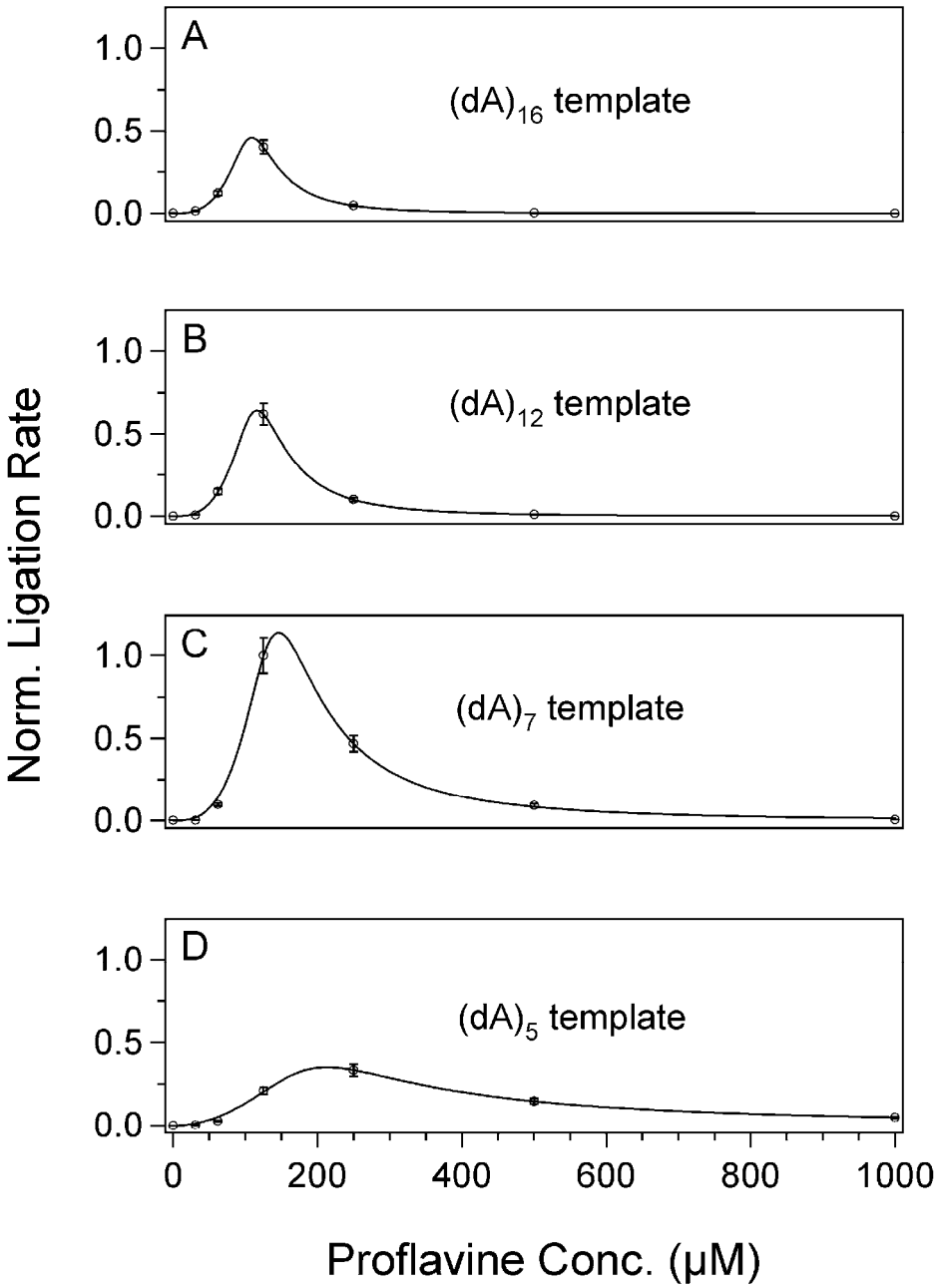


Figure 4.1. Proflavine-mediated ligation of substrates is affected by template $(dA)_n$ length. Plots of ligation rate for the formation of $(dT)_7$ product as a function of proflavine concentration for templates dA_{16}, dA_{12}, dA_7, and dA_5 (from A to D), respectively. Substrate, (3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$, and template strands were 7.0 μM in total dA and dT base concentration. The vertical 'ligation rate' scale is the same for all samples.

The ligation-rate profile observed for the template $(dA)_5$ has some distinct features from those of the three longer templates (Figure 4.1). Maximum ligation yield is reached at approximately 250 μM proflavine, which is approximately twice the concentration at which maximum ligation rate was observed for the three longer templates. The increase in ligation rate for the $(dA)_5$ template is also followed by a gradual decline from 250 – 1000 μM proflavine.

Length dependent changes in ligation can be explained by taking a closer look at the ligation system. The different behavior of reactions containing the $(dA)_5$ template may be due to the fact that this template is not long enough to accommodate the two substrates (i.e. dT_3 and dT_4) with full base pairing. This length effect is most apparent in a comparison of the ligation profiles of $(dA)_5$ and $(dA)_7$, which differ by only two base pairs, but only the latter can accommodate both substrates (Figure 4.1).

On the other hand, template strands $(dA)_{12}$ and $(dA)_{16}$, are longer than required to accommodate $(dT)_3$ and $(dT)_4$ substrates and therefore, present multiple sites for binding of the substrates. This type of substrate binding may keep the reactive ends of the two substrates distant from each other and could lead the declines in ligation profiles observed here. However, we note that this observed decline in product formation with increasing template length is counter to previous reports of template-directed synthesis [136]. It appears that in previous studies, conducted in the absence of an intercalator, longer templates were more efficient due to greater base preorganization that was provided by intrastrand stacking.

4.3.2. Mismatch Pairs and Their Effect on Ligation Activity

It is well known that mismatched base pairs substantially decrease the stability of a Watson-Crick duplex [137]. The binding of proflavine to nucleic acid duplexes is favored, in part, by the close match between the planar surface shapes of proflavine and the Watson-Crick base pairs. Thus, it is possible that mismatched base pairs could also affect intercalation-mediated synthesis. The extent to which mismatched base pairs affect formation of the ligation active complex is important because selectivity against mismatched pairings is essential for high fidelity information transfer. We have carried out rigorous experiments to determine if proflavine intercalation suppresses or enhances the expected deleterious effect of mismatched base pairs on proflavine-catalyzed ligation. For these experiments mismatches were introduced by using the same $(dT)_3$ and $(dT)_4$ oligonucleotide substrates, but with 7-mer oligonucleotides of various sequences as template stands. All ligation experiments were carried out at 4ºC in the presence of 100 μM proflavine.

Normalized ligation rates are presented in Figure 4.2 for twenty-eight different template sequences. Among these template sequences, $(dA)_7$, the only template that provides complete Watson-Crick pairing with substrates, exhibited the highest ligation rate (normalized to 1.0). This result is consistent with the Watson-Crick base pairs being the most stable pairing in a DNA duplex. The second most efficient template was d(AAGAAAA), with a normalized ligation rate of 0.4. The single G residue of this template would be expected to form a G·T mismatch base pair with the $(dT)_4$ substrate in the ligation active complex (Figure 4.2).

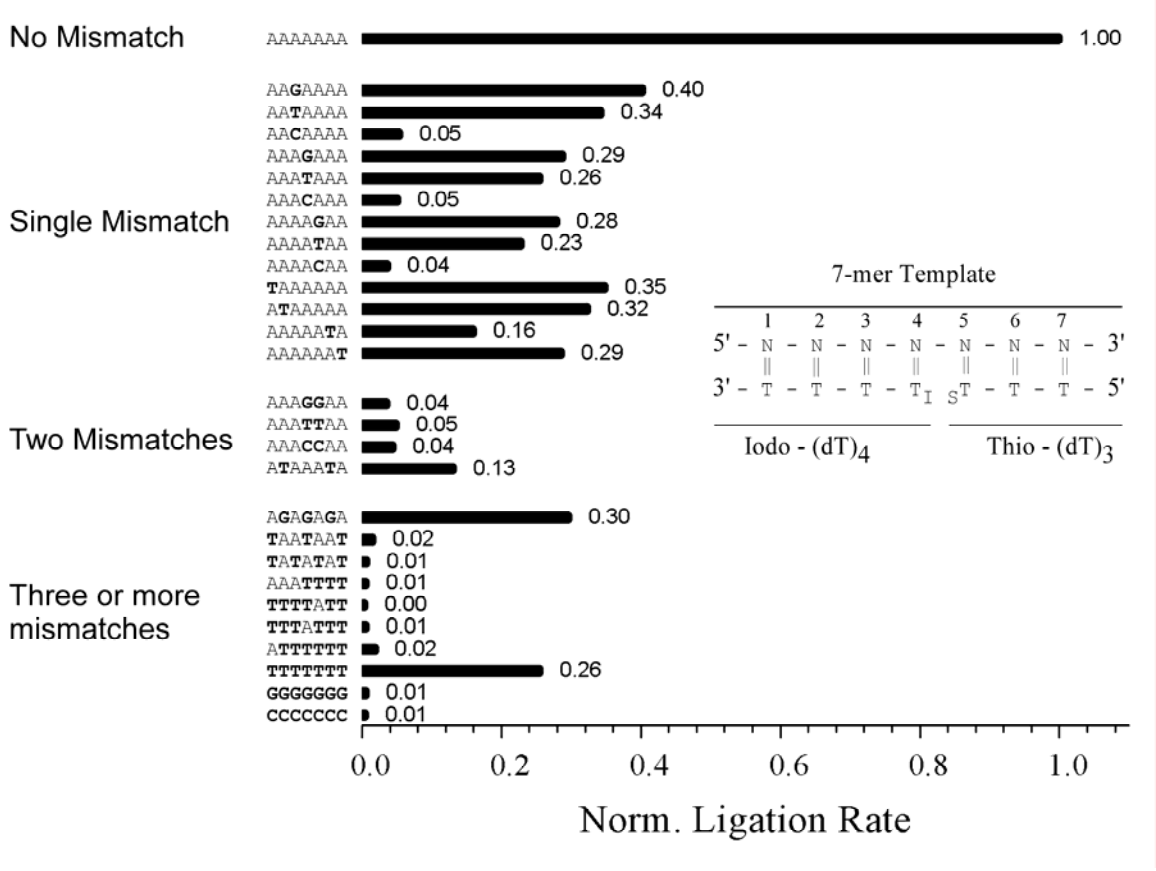


Figure 4.2. Mismatch base pairing and its effects on the ligation test system.
Normalized ligation rates of twenty eight different template sequences are shown in a bar graph. Samples contained 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ substrates, 100 μM proflavine and 7-mer template strand (5'→ 3') as indicated in the figure. Mismatched base pairs were strategically incorporated within the duplex structure by changing the bases within the template sequence (inset). Each data point represents an average intensity of three gel bands corresponding to the $(dT)_7$ product.

It has long been appreciated that G·T mismatches are similar in structure to the Watson-Crick base pairs [138], and can be accommodated within a DNA duplex with minimal alteration of the B-form helix structure [139]. Furthermore, thermodynamic studies have also shown that the G·T mismatch is the least destabilizing non-Watson-Crick base pair in a DNA double helix, although the stability of this base pair is still lower than the Watson-Crick base pairs [140]. Thus, it is understandable that d(AAGAAAA) would prove to be the second most efficient template, second only to $(dA)_7$. T·T and C·T base pair mismatches are more destabilizing that G·T base pairs in duplex DNA, with the relative order of stability being G·T > T·T >> C·T [141, 142]. This trend in mismatched base pair stability is qualitatively reflected in the lower normalized ligation rates for the templates d(AATAAAA) and d(AACAAAA) of 0.34 and 0.05, respectively.

The specific position of a single base pair mismatch within the ligation complex also has an appreciable effect on ligation rate. For example, the template d(AAGAAAA) exhibits a normalized ligation rate of 0.4, whereas the templates d(AAAGAAA) and d(AAAAGAA) exhibit normalized ligation rates of 0.29 and 0.28, respectively. The further reduction in ligation rates for the latter two templates is potentially due to the G·T mismatch being adjacent to the point of ligation (Figure 4.2). The same position-dependent trend in reduced ligation rates is also observed for the templates with single T·T mismatches, with normalized ligation rates of 0.34 for d(AATAAAA), 0.26 for d(AAATAAA) and 0.23 for d(AAAATAA). Templates with single C·T mismatches exhibit such low rates that differences between the positional effects of single C·T mismatches are within the limits of experimental error, with normalized ligation rates of

approximately 0.05 for d(AACAAAA), d(AAACAAA) and d(AAAACAA).

The decrease in ligation rates for G·T, T·T and C·T mismatches exhibit the same trend as the relative stability of these mismatches in a DNA duplex. However, the ligation rates for the single mismatch templates are larger than would be predicted based on the energetic penalty of each single base pair mismatch. Kallenbach's research has shown that the deleterious effects of having mismatches in an RNA duplex are compensated upon ethidium bromide intercalation [143]. In their studies with poly(rI)•poly(rC) duplexes, melting temperature of the duplex changes from 61.4ºC to 62.0ºC upon intercalation by EtBr. Upon introduction of mismatches in 10% of the duplex structure, melting temperature drops to 52.5ºC. However, upon ethidium intercalation of the mismatched duplex, duplex stability increases to 61.0ºC which is very similar to the duplex stability without mismatches. These results show that intercalation of nucleic acid structures by small molecules leads to tolerance of mismatches.

The template strands d(AAAGGAA), d(AAATTAA) and d(AAACCAA), which are each expected to introduce two tandem mismatches in the ligation-active complex, exhibited normalized ligation rates of 0.05 or less. These greatly reduced ligation rates are likely the result of both a reduced affinity of substrates for these templates as well as the disruption of local helical structure at the site of substrate ligation. This latter effect is suggested by the observation that the two T·T mismatches introduced by the template d(ATAAATA) only reduce the normalized ligation rate to 0.13, which is approximately three times greater than the rate observed for d(AAATTAA).

The most unexpected result of the twenty-eight templates studied was the relatively high normalized ligation rate of 0.26 exhibited by d(TTTTTTT). In contrast,

the oligonucleotides d(GGGGGGG) and d(CCCCCCC) exhibit ligation activities of approximately fifty-fold less which rules out the idea that somehow, homopurine or hompyrimidine templates are responsible for appreciable ligation activities. Template sequence d(GGGGGGG) may form a G-quartet and therefore may not function as a suitable template. However, no such problems exist for the d(CCCCCCC) template. Therefore, the activity of the d(TTTTTTT) template is particularly interesting when one considers that the normalized ligation rates observed for d(ATTTTTT), d(TTTATTT) and d(TTTTATT) are all 0.02 or less. The substantial difference between the ligation rates of d(TTTTTTT) and the three T-rich templates containing a single adenine base suggests that the ligation-active complex produced by d(TTTTTTT) in the presence of proflavine is fundamentally different from that of the other twenty-six templates containing mismatched base pairs. We propose that the d(TTTTTTT) template forms a intercalated double helix with substrates $(dT)_3$ and $(dT)_4$ with T·T base pairs, which may be the same T·T base pairs previously reported to exist in several nucleic acid structures [144-147].

We have used circular dichroism spectroscopy in order to explore the possibility whether a d(TTTTTTT) template can act as an efficient template in a ligation reaction. Addition of proflavine to a $d(T)_7$ strand leads to a large negative induced CD band at 450 nm and a substantial increase in the DNA band at 270 nm (Figure 4.3). Achiral proflavine seems to be interacting with a chiral $d(T)_7$ strand which leads to the appearance of large induced circular dichroism bands at 470 and 295 nm. These results point towards the possibility that substrate strands may organize alongside a $(dT)_7$ strand which could lead to ligation product formation. However, we are unable to fully understand whether

organization of substrates in the presence of a $(dT)_7$ template occurs in a unimolecular or biomolecular complex. It may be possible that the assembled state containing substrates and $(dT)_7$ template represents a kinetic trap and not the thermodynamically stable complex.

Our results with mismatch duplex pairs show that small molecule mediated template-directed synthesis is not strictly limited to Watson-Crick base pairing. The possibility exists where a prebiotic genetic system could have relied upon small molecule mediated assembly to promote non-canonical or even non-duplex secondary structures for information storage and transfer.

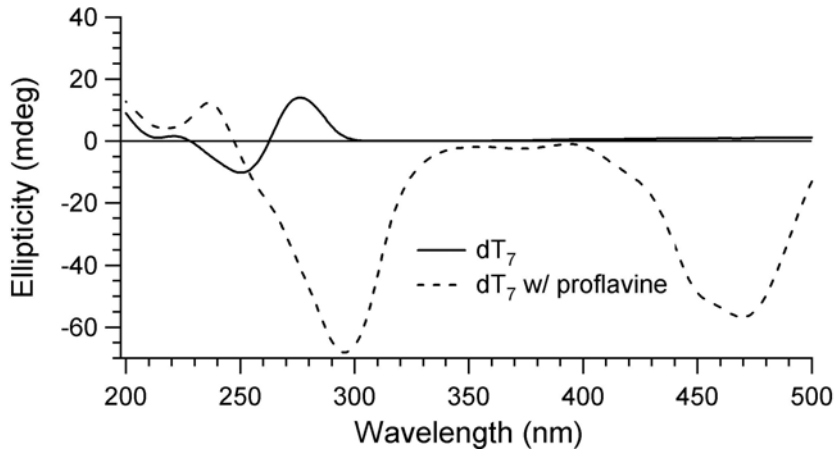


Figure 4.3. CD spectra of $(dT)_7$ strands with (dashed) and without (solid) proflavine. Sample: 480 μM in base, 120 μM proflavine, 10 mM Tris (pH 8.0), 4ºC.

4.3.3. An In-Depth Analysis of $(dT)_7$-Templated Ligation

4.3.3.1. Ligation Profiles of a T•T System

Mismatch studies have illustrated that strict Watson-Crick pairing is not necessary for small molecule mediated assembly of a ligation active complex. The most intriguing

result to come out of these studies was the finding that a T-rich template can yield appreciable ligation product in the presence of proflavine (Figure 4.2). We have initiated studies to explore whether this system behaves similarly to our Watson-Crick ligation test system. Ligation profiles as a function of proflavine concentration were carried for ligation complex concentrations ranging from 1 μM in base pair to 2000 μM in base pair. Figure 4.4 shows ligation profiles as a function of proflavine concentration.

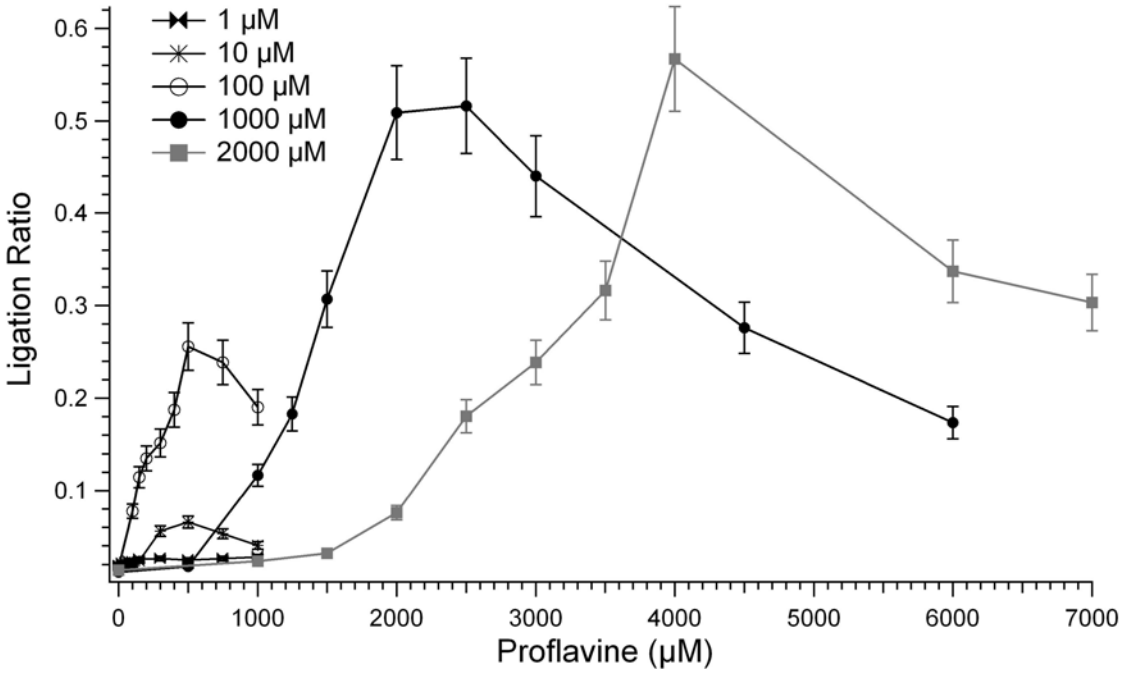


Figure 4.4 Ligation profiles as a function of proflavine concentration for a T•T system. All samples contained 100 mM 2-mercaptoethanol, 10 mM Tris (pH 8.0). Samples were incubated at 4ºC for 5 hours. Ligation complex concentrations (bp) are indicated in the figure caption. All samples were spiked with 0.1 μM in base of ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$ and the remaining amount is the chemically identical cold 5'-p-$(dT)_3$-ps-3' substrate. Ligation reactions where the radiolabeled amount is always kept constant are a direct reporter of enhancement or suppression of ligation activity due to variations in the ligation complex concentration.

The proflavine-catalyzed ligation of the $(dT)_3$ and $(dT)_4$ substrates in the presence of a $(dT)_7$ template is initially cooperative in proflavine concentration, however, at higher proflavine concentrations there is a substantial decline in ligation. Starting from 1 μM to 10 μM and up to 100 μM bp concentration, we observe a proportional increase in ligation activity (Figure 4.4). We note that the ligation activity seems to be limited by the concentration of substrate and template strands. However, further increase in ligation complex concentration from 100 μM to 1000 μM does not lead to a 10-fold increase in maximum ligation activity. Additional increase in ligation complex concentration to 2000 μM only leads to a small increase in maximum ligation activity. It may be possible that ligation activities at high concentrations of ligation complex are limited by proflavine. We observe aggregation and precipitation of proflavine at high concentration and this effect may be playing a role in suppressing ligation.

We note a correlation between maximum ligation activity (for each set of ligation complex concentration) and proflavine concentration. This systematic shift to higher proflavine concentration as ligation complex concentrations are increased suggests that ligation activity is governed by a loading mechanism. For example, the 1000 μM ligation complex has a concentration of nucleosides that is 2000 μM. We observe maximum ligation at ~ 2000 μM proflavine in this system. Similarly, the 2000 μM ligation complex has a nucleoside concentration of 4000 μM and maximum ligation occurs at 4000 μM proflavine.

4.3.3.2. Kinetic Studies of a T•T System

We wanted to determine initial rates of product formation for the T•T system. It is important to determine the shape of rate curves because deviations from linearity can suggest that ligation rates over the time period of the reaction are either slowing or increasing in their velocity. We have observed that higher concentrations of proflavine can cause suppression of ligation activity by a precipitation or aggregation phenomenon. Figure 4.5 is a rate profile for different ligation complex concentrations at a proflavine concentration where maximum activity is observed (Figure 4.4).

Ligation experiments where samples were taken at several different points during a 90-minute time period show linear behavior for all ligation complex concentrations except for 2000 μM. Linearity is observed only up to 60 minutes in this case. The 90-minute time point displays an unexpected increase in ligation and an appreciable deviation from linear behavior. We are unable to fully account for the significant differences in ligation rates. Room temperature ligation profiles as well as kinetic studies, similar to the Watson-Crick system in Chapter 4, need to be carried out for a direct comparison of the two systems.

An unusual type of helical structure may be forming in a ligation complex composed of thymine-thymine base pairs. There have been several occurrences of T•T base pairs in duplex as well as dimeric quadruplex structures [145, 146]. High resolution studies of a proflavine complex with T•T base pairs are necessary to fully understand this system.

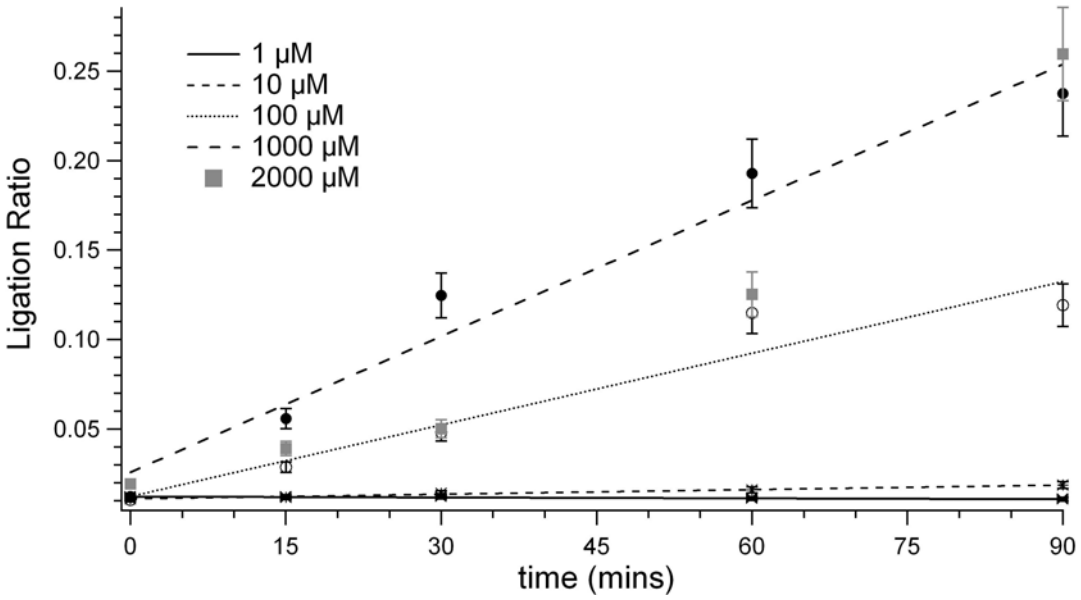


Figure 4.5. Kinetic studies of a proflavine catalyzed T•T system. All samples contained 10 mM Tris (pH 8.0) and 100 mM 2-mercaptoethanol. Ligation complex concentrations are indicated in the figure caption. Proflavine concentrations were 100 µM (1 µM bp), 400 µM (10 µM bp), 500 µM (100 µM bp), 2000 µM (1000 µM bp) and 4000 µM (2000 µM bp). 10 µl aliquots were flash frozen in liquid nitrogen at 0, 15, 30, 60, 90 minutes. Samples were freeze dried in a lyophilizer. Experiments were done at 4ºC. Data points are an average of independent gel bands in duplicate.

4.3.3.3. Structure of a T•T Mismatch

Thymine-thymine base pairs have been previously reported to occur in nucleic acid structures [144-147]. Buck and coworkers have shown that parallel thymine-thymine mismatches can form in aqueous solution [146]. Figure 4.6 illustrates one of the potential structures of a thymine-thymine mismatch.

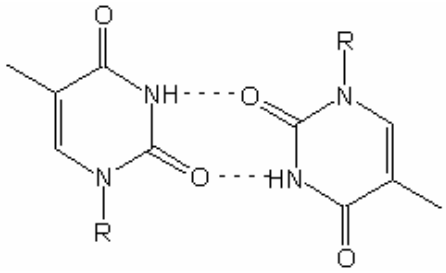


Figure 4.6. Structure of a thymine-thymine mismatch [146].

4.3.4. Effect of Midwife Structure and Binding Mode on Assembly of a Ligation Active Complex

We originally selected proflavine as a possible midwife molecule to enhance template-directed synthesis based upon three criteria: 1) proflavine binds to nucleic acid duplexes with micromolar affinity; 2) proflavine interacts with DNA by base pair intercalation; and 3) the geometric structure of proflavine is an excellent match with the Watson-Crick base pairs. As a means to qualitatively compare the relative importance of these three properties for a molecular midwife, the ability for several other DNA-binding molecules to act as midwives was also tested using the ligation assay with the $(dT)_3$, (dT_4) substrates and $(dA)_7$ as the template. The small molecules used in these experiments, shown in Figure 4.7A, were chosen from three categories: duplex DNA intercalators (ethidium, daunomycin, actinomycin, acridine orange, proflavine), triplex DNA intercalators (coralyne, BePI), and minor groove binders (distamycin, DAPI).

All of the small molecules represented in this study, except acridine orange, reportedly bind AT-rich DNA duplexes with equilibrium binding affinities that are within one order of magnitude of one another [72]. In Figure 4.7B normalized ligation rates are presented for reactions carried out in the presence of each of these ligands at a concentration of 100 μM.

Acridine orange (AO) is a close structural analogue of proflavine with a similar binding constant to DNA [148]. However, AO exhibits a 10-fold decline in the normalized ligation rate (0.12) with respect to proflavine. The decline in ligation rates in the presence of AO with respect to proflavine can may be partially due to the large dimerization constant of AO, which is an order of magnitude higher than proflavine [149,

150]. This greater propensity for AO to self-associate may alter its interaction with DNA at higher concentrations. It is also possible that the presence of four exocyclic methyl groups in the structure of acridine orange inhibits the chemical step of the ligation reactions. Actinomycin, daunomycin and ethidium are other members of duplex intercalating class of molecules that show minimal ligation of $(dT)_3$ and $(dT)_4$ substrates in the presence of a $(dA)_7$ template (4.7B). Unlike proflavine, these small molecules each contain bulky side chains that are external to the planar ring system expected to intercalate DNA base pairs. Actinomycin is distinct from the other intercalators tested, as this molecule is known to preferentially bind GC-rich sequences over AT-rich sequences [72]. As would be predicted for a ligation test system containing only A·T base pairs, actinomycin exhibits minimal enhancement of the ligation reaction over background. Ethidium and daunomycin, on the other hand, are known to bind AT-rich duplexes with binding affinities comparable to proflavine. Both these molecule contain a positive charge similar to proflavine. Thus, the apparent inability for ethidium and daunomycin to enhance ligation suggests that a close structural match with the desired pairing geometry and/or complete planarity of the small molecule is an important requirement for midwife activity in a ligation reaction.

BePI and coralyne (4.7A) are small molecules that also intercalate DNA, but preferentially bind triplex over duplex DNA [110, 112]. Normalized template-directed ligation rates in the presence of BePI and coralyne were determined to be 0.04 and 0.02, respectively. BePI and coralyne bind to AT-rich duplex DNA with equilibrium constants that are within an order of magnitude of that measured for proflavine [72](J. Chaires, personal communication).

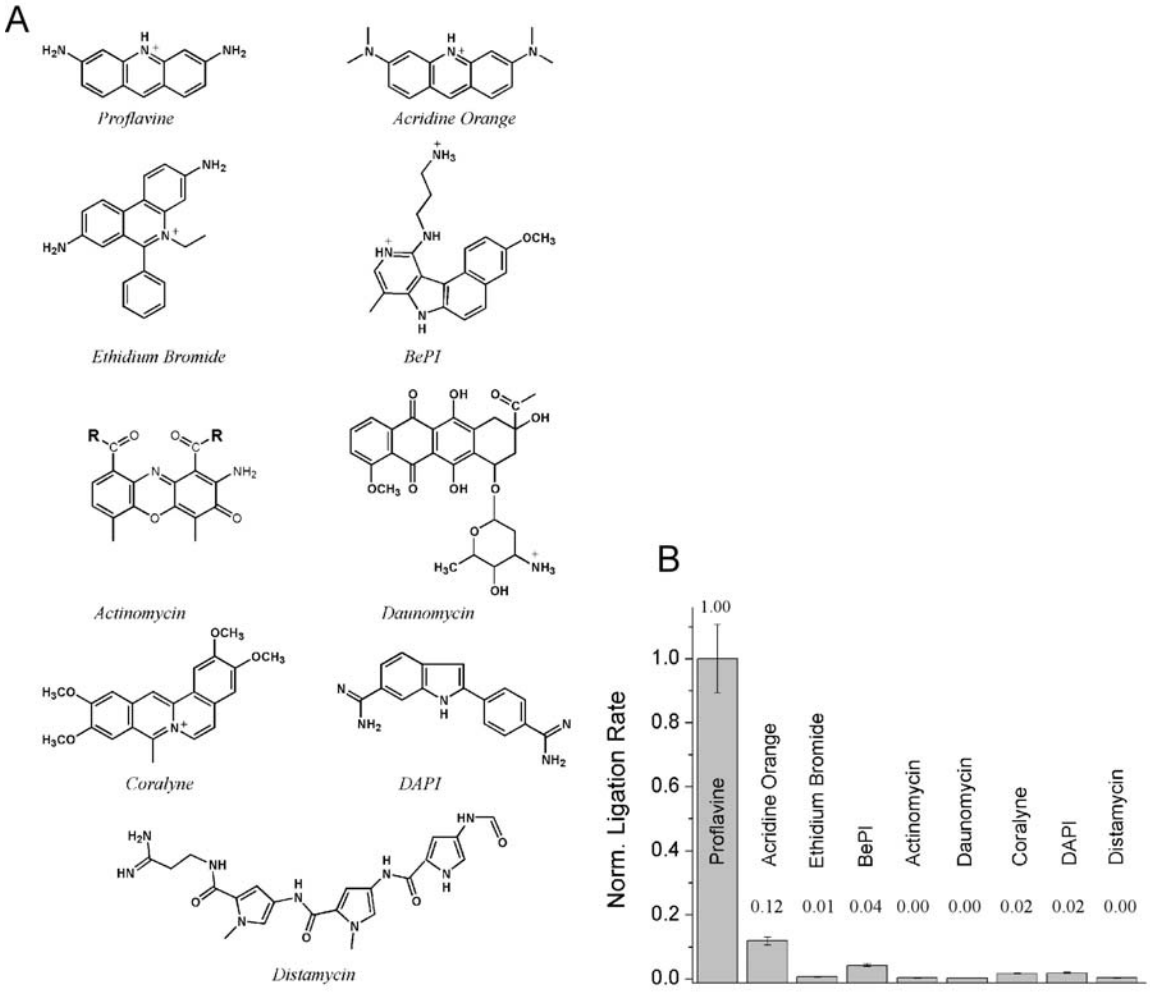


Figure 4.7. Effect of small molecule structure on template-directed ligation reactions. (A) Chemical structures of small molecules used in the study. **R** group in the actinomycin structure represents a cyclic peptide chain. (B) Bar graph illustrates normalized ligation rates in the presence of different small molecules. Samples contained 3'-phosphorothioate-$(dT)_3$, 5'-iodo-$(dT)_4$ substrates, $(dA)_7$ template strand, and 100 μM ligand. Error bars represent standard deviations of values from independent assays performed in triplicate.

Thus, reduced binding affinity cannot fully account for the observed decrease in ligation rates for these two molecules in comparison to proflavine, and again suggest that optimum ligation is achieved when the geometric structure of an intercalating molecule is a close match to that of the desired base pairing structure.

Distamycin and DAPI are uncharged, minor groove binding agents with approximately 3-fold higher affinities for AT-rich duplex DNA than proflavine [72, 151]. Distamycin binds in the minor groove along a binding site that spans 4-5 base pairs in a 1:1 or 2:1 (Distamycin:DNA) ratio [151, 152], whereas DAPI binds in a 1:1 ratio with a binding site consisting of 3 base pairs [151, 153]. Distamycin has been shown to enhance the chemical ligation of plasmid DNA with sticky ends [154]. However, our data revealed essentially no enhancement of ligation by distamycin (normalized rate <0.01) or DAPI (normalized rate of 0.02) (4.7B). One important difference between our ligation test system and plasmid ligation is that the former is a trimolecular DNA assembly, whereas the latter is a bimolecular DNA assembly. Additionally, the extended duplexes of a plasmid that flank the ligation site may provide a degree of preorganization that is necessary for a minor groove binding molecule to bind. It has previously been demonstrated that a single nick in one strand of a duplex interrupts minor groove binding sites for molecules such as DAPI [153], which is indicative of the fact that minor groove binding molecules require a minimal level of stability in minor groove structure. In contrast, proflavine may be able to assemble the ligation-active complex that contains a break in one backbone due to the fact that intercalation is a much shorter-range interaction (i.e. only between stacked base pairs).

4.3.5. Increased Ionic Strength Inhibits Proflavine Mediated Ligation

Orgel and co-workers have demonstrated an absolute necessity for molar concentrations of NaCl or $MgCl_2$ to achieve template-directed synthesis with methylimidazole activated nucleotides [136]. The proflavine-mediated ligation reactions described above contained 10 mM NaCl. In order to explore the sensitivity of proflavine-mediated ligation to changes in ionic strength, we carried out experiments designed to determine the optimal salt concentration for proflavine-mediated ligation. The normalized rates of ligation reactions carried out in the presence of 100 μM proflavine as a function of ionic strength are shown in Figure 4.8A. Product formation was observed to decrease exponentially as a function of increasing NaCl concentration. A similar trend is also observed when $MgCl_2$ is used as the counter ion species. A comparison of the NaCl and $MgCl_2$ ligation data plotted as functions of ionic strength (Figure 4.8A) demonstrate that $MgCl_2$ reduces proflavine-mediated ligation to a greater extent than NaCl at all comparable ionic strengths tested.

The binding affinity of proflavine to DNA is known to decrease with increasing salt concentration [155]. Thus, one might expect that the observed decrease in ligation rate with increasing salt concentrations results from a decrease in proflavine binding to the ligation active complex.

Optimum ligation in our ligation test system was observed at ~100 μM proflavine, which is above the concentration required for full intercalation of the ligation active complex by proflavine. In addition to intercalative binding, acridine dyes are known to stack alongside the phosphate backbone of nucleic acids at higher dye:DNA base pair concentrations than 1:2 [92, 156].

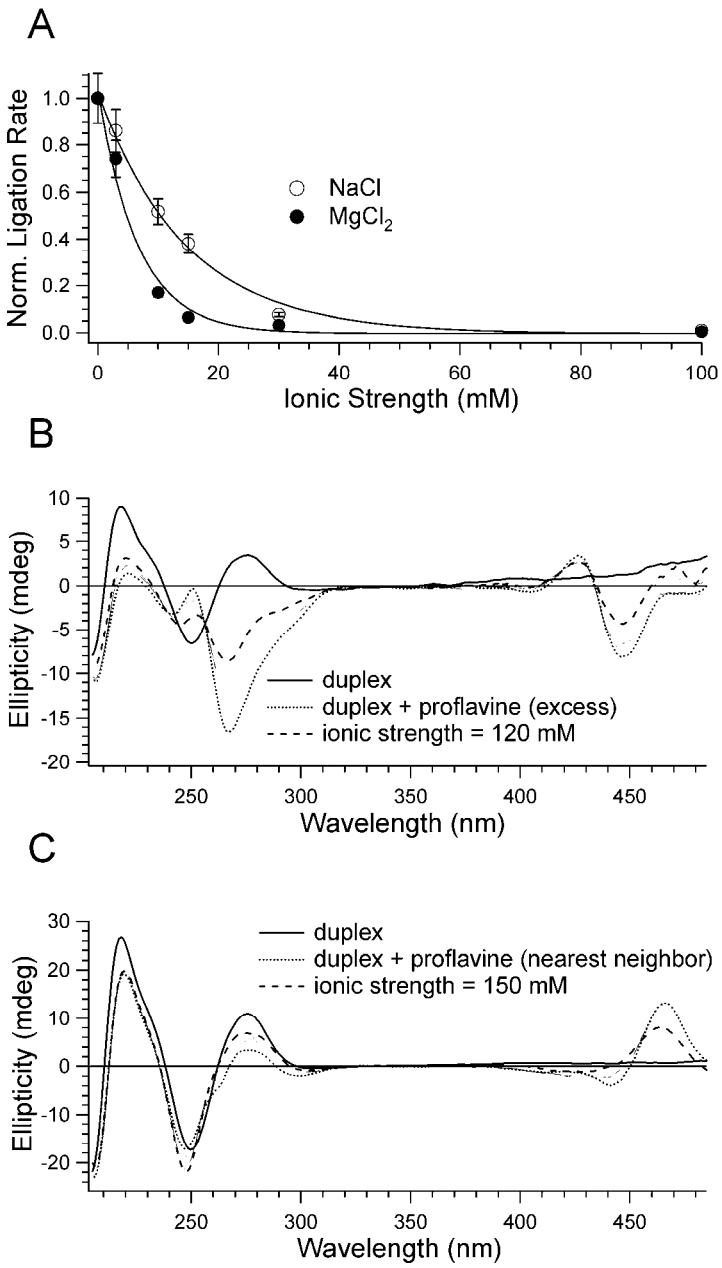
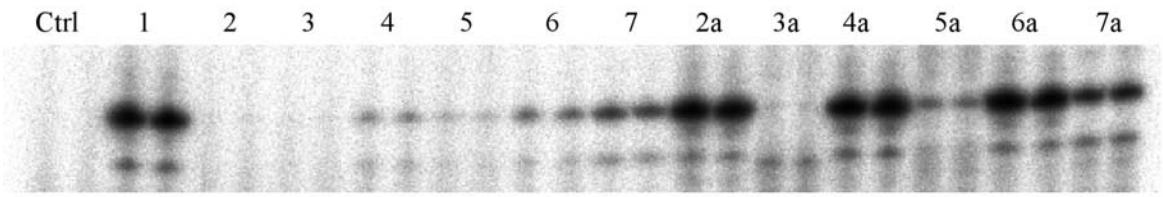


Figure 4.8. Proflavine-mediated ligation is dependent on solution ionic strength.
(A) Samples contained 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ substrates, $(dA)_7$ template strand and 100 μM proflavine. Note that all samples contained 10 mM Tris buffer which contributes to the overall ionic strength of the sample solution. Error bars reflect standard deviations of values from independent assays performed in triplicate. (B) CD spectra illustrating the effect of ionic strength on the binding of proflavine to a duplex $(dA)_7$:$(dT)_7$ sample. Proflavine concentration was in excess to that required by nearest neighbor intercalative binding and is similar in concentration to the ligation reaction. (C) CD spectra showing the effect of ionic strength on nearest neighbor proflavine binding to a $(dA)_7$:$(dT)_7$ sample.

As a means to investigate if increasing ionic strength suppresses the binding of proflavine to the ligation active complex in either mode of interaction, we have investigated the effect of changing ionic strength by monitoring the circular dichroism spectra for proflavine binding to (dA_7):(dT_7) duplex DNA under two different conditions: proflavine bound in excess (Figure 4.8B) and proflavine bound at a concentration allowing nearest neighbor intercalative binding (Figure 4.8C). With increasing ionic strength, the CD spectra suggest a decline in proflavine binding to the duplex structure, as indicated by decreases in peak intensities at around 270 nm and 450 nm. However, the normalized ligation rate (Figure 4.8A) drops to a negligible level at an ionic strength value where a strong induced CD signal for proflavine binding to DNA is still observed (Figure 4.8B, 4.8C). Thus, the CD spectra indicate that proflavine binding to a ligation-active complex persists past the salt concentration at which ligation is no longer detected. Nevertheless, it remains a possibility that with increasing salt concentration the structure of the proflavine:DNA ligation-active complex is perturbed in a manner that adversely affects the coupling rate of the reaction.

4.3.6. Presence of Crowding Agents Affects Ligation Activity

The presence of crowding agents on small molecule mediated ligation of nucleic acids was investigated by adding a crowding agent to the sample solutions. Three different types of PEG molecules (PEG 400, 1500, 6000) were added to the ligation mixture at two different percentages each (0.5% and 4.0%). Denaturing PAGE analysis in Figure 4.9 illustrates the effect of crowding agents on small molecule mediated ligation of nucleic acids in the previously described test system.



2. PEG 400 (0.5 %) 3. PEG 400 (5 %)

4. PEG 1500 (0.5 %) 5. PEG 1500 (5 %)

6. PEG 6000 (0.5 %) 7. PEG 6000 (5 %)

Figure 4.9. Polyethylene Glycol (PEG) significantly affects ligation activity. Denaturing polyacrylamide gel electrophoresis analysis illustrates the effect of PEG on the ligation yield of 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ using $(dA)_7$ as a template strand. Ctrl: Only ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$ as blank correction. All samples were run in duplicate. Lane 1: Standard with 100 μM proflavine. All other lanes ontain PEG (type and percentage is indicated above). Lanes 2 – 7 contain no proflavine whereas lanes 2a – 7a control 100 μM proflavine.

Lanes 4-7 in Figure 4.9 show that polyethylene glycol catalyzes the ligation of substrates even in the absence of proflavine. This intriguing result highlights the importance of a cellular environment and its effect on small molecule-mediated assembly and ligation of nucleic acids. Highest ligation activity in the absence of proflavine is observed using PEG 6000 at a solution concentration of 5.0%. PEG 1500 at a 5% content is much less favorable in the coupling of substrates. It will be interesting to observe whether polyethylene glycol mediated reaction is possible in the absence of $(dA)_7$ template strand.

Lanes 2a – 7a illustrate the effect of PEG on proflavine-mediated assembly of nucleic acids. Minimal change in ligation activity is observed at a 0.5% solution concentration of polyethylene glycol (2a, 4a, and 6a) compared to a standard without PEG (lane 1). However, lanes 3a, 5a, and 7a show that PEG at a 5% concentration leads

to an appreciable decline in ligation activity. It is interesting to note that the deleterious effects of having 5.0% crowding agent are minimized as the size of PEG increases from PEG 400 to PEG 6000.

The crowded cellular environment contains a myriad of components including but not limited to organelles, proteins, nucleic acids, lipids, carbohydrates, ions etc. Ligation assays in the presence of PEG as a crowding agent show that solution conditions must be taken into considerations in the design of experiments where assembly and ligation of nucleic acids using small molecules is studied.

4.3.7. Proflavine Assembles Ligation Complexes with Hybrid Nucleic Acid Structures

Proflavine is conjugated, planar heterocyclic molecule having a structure that is expected to have minimal contacts with the backbone of nucleic acids. We wanted to investigate whether proflavine can assemble a stable ligation complex with RNA templates. For this analysis, the substrates were the same as used before: 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$. The templates used were $(dA)_7$, $(rA)_7$, and 2'-O - Me $(rA)_7$. In Figure 4.10, we present results from proflavine-catalyzed ligation of $(dT)_3$ and $(dT)_4$ on the DNA as well as the two variable RNA template strands.

Denaturing gel in Figure 4.10A illustrates the effect of backbone structure on the assembly of a proflavine-mediated ligation active complex. Proflavine mediated ligaiton is appreciably lower for the RNA templates in comparison to the $(dA)_7$ strand. In a direct comparison of the $(dA)_7$ and $(rA)_7$ templates, an increase in ligation followed by a decline in ligation activity at higher proflavine concentrations is observed (Figure 4.10B). The

profile for the 2'-O-methyl $(rA)_7$ template is significantly different than $(dA)_7$ and $(rA)_7$ template strands. We observe that ligation activity increases and then remains at a plateau before a decline in ligation is observed. Additionally, the shift in maximum ligation activity towards a higher proflavine concentration in comparison to the $(dA)_7$ and $(rA)_7$ templates suggests that proflavine binds with a lower affinity to a ligation complex with 2'-O-methyl $(rA)_7$ template which can, in turn, affect its ability in assembly of a ligation complex. It is not surprising to observe that proflavine assembles a ligation active complex with an $(rA)_7$ and 2'-O-methyl $(rA)_7$ strands given the fact that proflavine primarily interacts with nucleic acid bases by stacking and is expected to make minimal contacts with sugar or phosphate groups in the backbone of nucleic acids. Competition dialysis experiments have shown that proflavine can bind hybrid nucleic acid structures albeit with a reduced affinity in comparison to DNA:DNA structures [72] (J. Chaires, personal communication). It is possible that the assembly of a hybrid structure perturbs the positioning of the reactive groups and thereby leads to lower ligation activity. In any case, our data shows that small molecule-mediated assembly of nucleic acids can be expanded beyond the scope of DNA:DNA structures.

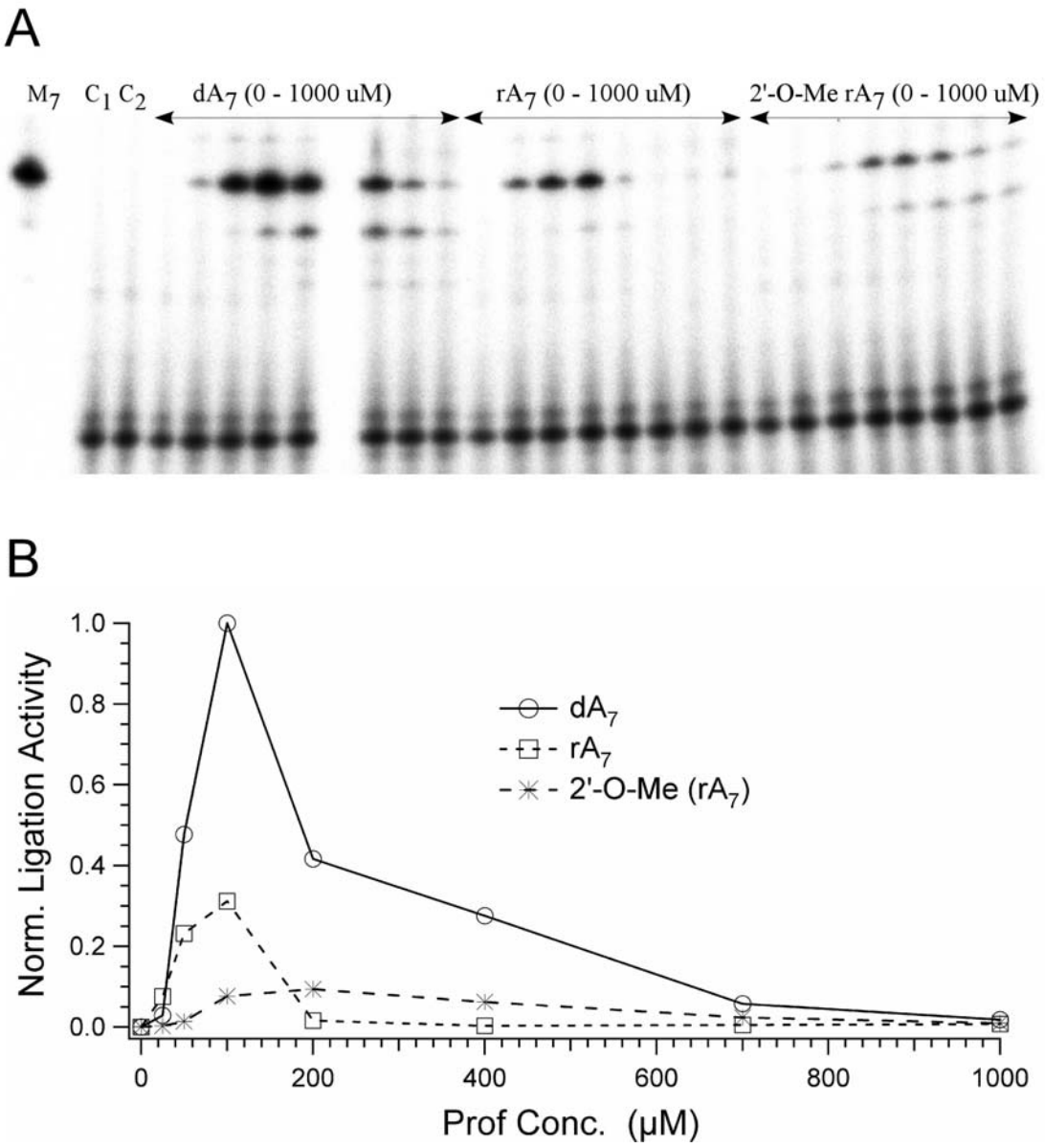


Figure 4.10. Gel electrophoresis of proflavine-mediated assembly involving hybrid complexes. (A) Denaturing polyacrylamide gel electrophoresis analysis that illustrates the effect of proflavine on the ligation yield of 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ using $(dA)_7$, $(rA)_7$, and 2'-O-methyl $(rA)_7$ as template strands. Lane C1: Only ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$. Lane C2: Substrates ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$. Concentrations were 1 μM in strand for both substrates and template strand. Proflavine concentration, from left to right for each template strand, is 0, 25, 50, 100, 200, 400, 700, and 1000 μM, respectively. All reaction mixtures were incubated for 24 h at 4ºC. Lane M_7: Molecular weight marker band of $(dT)_7$. (B) Plots of normalized ligation for formation of the $(dT)_7$ product as a function of template strands, $(dA)_7$, $(rA)_7$, and 2'-O-methyl $(rA)_7$. Rates shown have been normalized to the maximum of the fit of the data for proflavine with the DNA template.

4.4. CONCLUDING REMARKS

Intercalation-mediated assembly of nucleic acid strands represents a powerful approach to increasing the efficiency of protein-free template-directed synthesis. A small molecule that intercalates the bases of DNA and RNA can enhance the ligation rate of short substrate strands along a template strand by three orders of magnitude. A systematic study of several experimental parameters, including template length, small molecule structure, solution composition (ionic strength and presence of crowding agents) has provided valuable insights regarding future optimization of small molecule-mediated ligation reactions. Our results demonstrate that small molecule structure and mode of binding to nucleic acids are critical factors in the assembly of a ligation-active complex, not simply binding affinity. The finding that proflavine assembles hybrid ligation complex illustrates that ligation complex need not be limited to strict DNA:DNA or RNA:RNA structures. These results are very important in the context of nucleic acid origin as we show that small molecule mediated assembly can be used to include diverse array of nucleic acid structures.

The serendipitous discovery of a ligation-active complex containing exclusively T•T base pairs assembled in the presence of a small intercalating molecule demonstrates that intercalation-mediated template-directed synthesis is not strictly limited to Watson-Crick base pairing. Thus, artificial genetic systems based on pyrimidine-pyrimidine or purine-purine pairing may be possible with intercalation-mediated synthesis. Pairing of like bases on substrate and template strands could provide a more direct path to artificial self-replication than Watson-Crick base pairing, because product strands would be identical to template strands. Finally, the ability for a small intercalating molecule to

dictate base pair formation provides support for the previous proposal that small molecules in the pre-RNA World could have been responsible for originally selecting the Watson-Crick base pairs of present day nucleic acids.

CHAPTER 5

COMPLEX BINDING BEHAVIOR OF PROFLAVINE IN ASSEMBLY OF A LIGATION ACTIVE COMPLEX

5.1. INTRODUCTION

In the previous chapters, we provided evidence that small molecule intercalation can greatly enhance the coupling of oligonucleotide substrates in a template-directed reaction [85]. In these ligation reactions we also observed an appreciable decline in ligation activity at higher concentrations of proflavine. Experiments discussed in this chapter were designed to uncover the origin of this decline in ligation activity. Results presented here illustrate that it is not only necessary for a molecular midwife to promote formation of a ligation-active complex, but that other non-active ligation complexes must not be promoted by the midwife. We have found that the binding behavior of proflavine to a ligation active complex is more complicated than we previously anticipated. We also find that the assembly of a ligation active complex is dependent upon the additional binding modes uncovered here. We have used circular dichroism spectroscopy to study various aspects of proflavine-mediated assembly of a ligation complex. Binding mode, stoichiometry, binding affinity and stability of the complex can all be probed in detail by utilizing various CD spectroscopy techniques. These studies should facilitate future developments that more fully exploit the molecular midwife approach to template-directed ligation and replication of nucleic acids.

5.2. EXPERIMENTAL PROCECURES

5.2.1. Preparation of Nucleic Acid and Proflavine Stock Samples

Substrate oligodeoxynucleotides were synthesized on an automated synthesizer using the phosphoramidite coupling chemistry. Synthesis of 3'-phosphorothioate-$(dT)_3$ and 3'-phosphorothioate-(dT) was accomplished using a 3'-phosphate controlled pore glass support (Glen Research) where the oxidation reagent normally added during the first nucleotide coupling cycle was replaced by a sulfurizing reagent (Glen Research). The 5'-iodo-$(dT)_4$ substrate oligo was synthesized using the commercially available iodination reagent (Sigma). Iodination reaction was carried out on a deprotected 5'-OH-$(dT)_4$ oligo on the solid CPG support according to published procedures [157]. 3'-phosphorothioate, 5'-phoshpate-$(dT)_3$ and 3'phosphate, 5'-phoshpate-$(dT)_3$ oligos were synthesized by carrying out phosphorylation and sulfurization steps according to standard artificial DNA synthesis protocols. Following cleavage and deprotection, substrate oligonucleotides were purified by HPLC and characterized using mass spectrometry. All other oligonucleotides (Integrated DNA Technologies) were desalted and used without further purification. Stock solutions of oligonucleotides were prepared by resuspending freeze-dried purified samples in deionized H_2O. Oligonucleotide concentrations were determined spectrophotometrically. Stock samples of proflavine were made in dH_2O by using proflavine hemisulfate salt (Sigma). Mild heating was required to dissolve proflavine in dH_2O. Concentrations were determined spectrophotometrically using the extinction coefficient, $\varepsilon_{445\ nm} = 38{,}900\ M^{-1}\ cm^{-1}$.

5.2.2. Template-Directed Ligation Assay

The 3'-phosphorothioate-$(dT)_3$ substrate was radioactively labeled with ^{32}P-phosphate at the 5'-end to a stock concentration of 1 μM in bp. Ligation reactions were carried out in 10 μl volumes in a solution containing 10 mM Tris buffer (pH 8.0) and 100 mM 2-mercaptoethanol. The substrate 5'-iodo-$(dT)_4$, the substrate 3'-phosphorothioate-$(dT)_3$, and the template $(dA)_7$ were each added to the reaction buffer (concentrations indicated in the figure caption). $(dT)_3$ substrate in the sample is comprised of 0.1 μM in base of 5'-^{32}phosphate-3'-phosphorothioate-$(dT)_3$ and the remaining amount is the chemically identical 5'-phosphate, 3'-phosphorothioate-$(dT)_3$ substrate. The presence of 2-mercaptoethanol in the reaction buffer was necessary to prevent dimerization of the 3'-phosphorothioate $(dT)_3$ substrate. All reaction mixtures were incubated at 22ºC for 1h. Ligation reactions were stopped by plunging reaction test tubes into liquid nitrogen and freeze-drying for 1h on a lyophilizer.

5.2.3. Ligation Analysis and Product Quantification

Freeze-dried reaction samples were resuspended in 10 μl of a dye solution (8 M urea) and loaded on to a denaturing 30% polyacrylamide gel (19:1::acrylamide:bisacrylamide). Gels were subject to electrophoresis at a constant power of 70 W for 5 hours. The relative yield of $(dT)_7$ product for each reaction was determined by imaging the gel on a Fuji Phosphor Imager and quantifying the integrated intensity of gel bands in each lane that corresponded to the $(dT)_7$ ligation product using the software package Image Gauge V3.12. Background correction was accomplished by subtracting from all reaction samples the integrated intensity of an area in a control lane run with only ^{32}P-end-labeled $(dT)_3$.

5.2.4. Circular Dichroism Spectroscopy

Circular dichroism experiments were performed using a JASCO J-810 CD spectropolarimeter equipped with a Peltier temperature control device. Spectra were acquired using 1 mm, 2 mm, 5 mm, 10 mm, 50 mm and 100 mm path length cells (Helma). Spectra are an average of 3 scans acquired at a scan speed of 500 nm/min. Samples contained 10 mM Tris buffer (pH 8.2) and DNA concentrations as indicated. We have carried out three types of analyses using circular dichroism spectroscopy.

5.2.4.1. Melting Temperature Determination

CD melting profiles were acquired using a 2 mm path length cell by increasing the sample temperature at a rate of 1.0°C/min from 4 to 40°C. Spectra were acquired at 2°C intervals. Concentration of ligation complex was 240 μM in bp. Proflavine concentration in the samples were 120 μM (to that allowed by nearest neighbor exclusion), 240 μM, 360 μM, and 480 μM.

5.2.4.2. Titration Analysis

The following procedure is an example in which titration analyses were carried out. A 1000 μl sample containing 10 mM Tris (pH 8.0) and 100 μM bp $(dA)_{16}$•$(dT)_{16}$ duplex was prepared. A titrant solution containing 10 mM Tris (pH 8.0), 5.0 mM proflavine, and 100 μM bp $(dA)_{16}$•$(dT)_{16}$ duplex was also prepared. To the DNA sample, titrant was added in specific increments so as to obtain a desired phosphate to proflavine (R) ratio. Presence of 100 μM bp $(dA)_{16}$•$(dT)_{16}$ in the titrant solution was necessary to avoid dilution of DNA during the course of the titration. Titration studies were performed in

two distinct ways: 1) proflavine was titrated into a DNA sample, and 2) DNA was titrated into a proflavine sample. Titration analyses were carried out at 4 and 22°C.

5.2.4.3. Dilution Studies

Binding constant determination was carried out at 22°C using CD cells with various path lengths. Duplex $(dA)_{16}$•$(dT)_{16}$ concentration was 200 μM in bp. Two samples having proflavine concentrations of 100 μM (R = 4, allowed by NNEP) and 400 μM (R = 1, fully loaded complex) were prepared. Samples were diluted with 10 mM Tris buffer (pH 8.0). Loss in CD signal upon dilution was compensated at certain points by changing to a CD cell with a longer path length. Similarly, binding constants were also determined for an unstable $(dA)_7$•$(dT)_7$ sample. Two samples, 250 μM proflavine (R = 4) and 1000 μM proflavine (R = 1), were prepared with the $(dA)_7$•$(dT)_7$ concentration 500 μM in bp for both sample.

5.3. RESULTS AND DISCUSSION

5.3.1. Assembly of a Ligation Active Complex

We have developed an experimental system to explore the ability for small molecule intercalation to promote the template-directed ligation of nucleic acids [85]. This test system utilizes modified substrates (3'-phosphorothioate dT_3 and 5'-iodo dT_4) and a $(dA)_7$ template strand. Intercalation of substrate and template strands produces a *ligation-active complex* in which the phosphorothioate and iodo groups of the substrates are sufficiently close to react and form a covalently-linked product that is an analogue of $(dT)_7$ (Figure 3.2). The small molecule proflavine (Figure 3.2B), known to intercalate Watson-Crick duplexes, can increase the ligation rate of oligonucleotides in our ligation test system by up to 1000-fold [85].

We note that in our previous studies, the 3'-phoshorothioate-$(dT)_3$ substrate upon radiolabeling at the 5'-end carries negative phosphate groups at both ends. The significant negative charge on a trimer substrate composed of thymine bases may affect equilibrium secondary structure formation (Figure 5.1). It has been shown by Turner and coworkers that a phosphate groups at the 5'end of an RNA tetramer (pGGCC) is stabilizing (ΔT_m = 3.3ºC) whereas the 3'-phosphate group (GGCCp) leads to a decrease in melting temperature (ΔT_m = - 1.5ºC) compared to the unphosphorylated oligonucleotide [158]. End-fraying effects in short $(dT)_n$•$(dA)_n$ oligomers may be reduced by the presence of phosphate groups which may play a capping role [159]. In our earlier experiments, PNK radiolabeling only added phosphate groups to the 5'-end of approximately 0.1% of total substrate strands due to practical limitations. This leads to solution conditions where the reporter strands have phosphate groups at both 3'- and 5'-

ends whereas 99.9% of the substrates carry a phosphate group at only their 3'-end. Substrates strands with differential charge density may quantitatively affect the formation of the ligation active complex. To avoid the complications due to differential charges, we have undertaken chemical synthesis of phosphorothioate substrates where the 3'-end contains the phosphorothioate group and the 5'-end contains a radiostable ("cold") phosphate group. Ligation experiments have been carried out in a manner where a fraction of substrate is radiolabeleld and the remainder is the chemically identical cold substrate (Experimental Section). Spectroscopy experiments have also been carried out where the solution conditions parallel those in the ligation experiments.

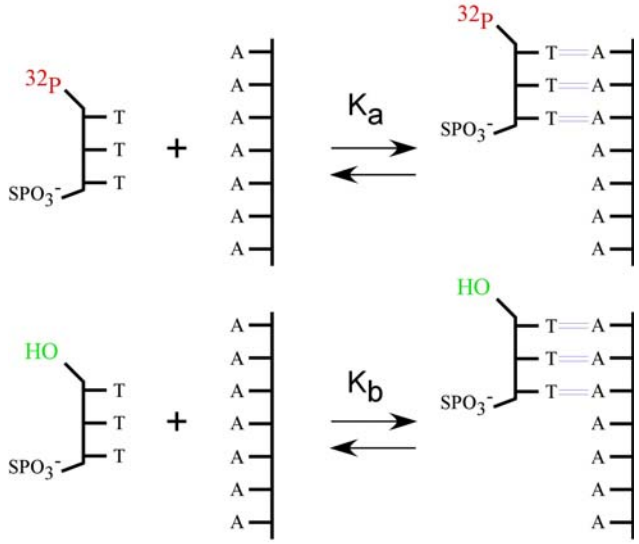


Figure 5.1. Schematic representation of equilibrium duplex formation with terminal phosphates. Equilibrium duplex formation (K_a) for a substrate with phosphates at both termini can be appreciably different than a substrate with phosphate group at only one end (K_b). Charge-charge repulsions as well as a change in end-fraying effects due to the presence of phosphate groups may play a role in differential equilibrium duplex formation.

A representative gel image shown in Figure 5.2 illustrates product formation as a function of proflavine concentration in a template-directed ligation reaction. Gel electrophoresis data shows that $(dT)_7$ product formation is undetectable for a reaction containing $(dT)_3$, $(dT)_4$ substrates and $(dA)_7$ template strand. Addition of proflavine to the reaction mixture promotes product formation and the intensity of the $(dT)_7$ band increases as a function of proflavine concentration. At higher proflavine concentrations, we again observe a decline in product formation.

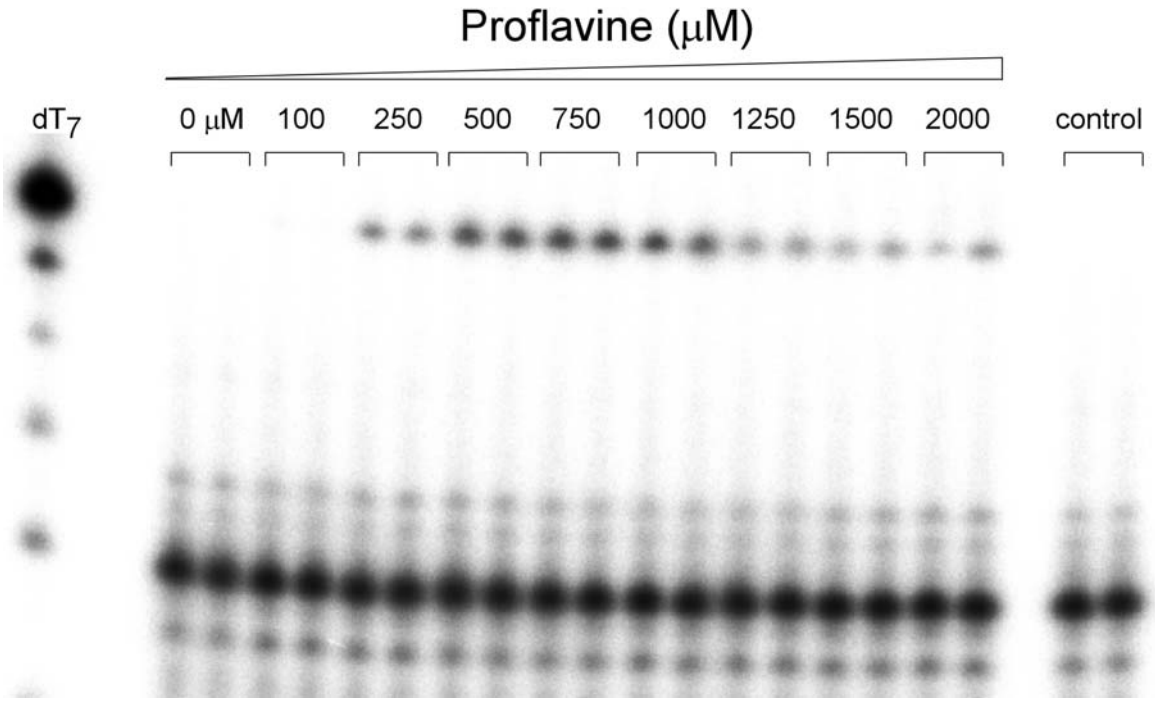


Figure 5.2. Denaturing PAGE analysis illustrating proflavine-mediated assembly. Effect of proflavine on the ligation yield of 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ using $(dA)_7$ as a template strand is illustrated. Control: Only ^{32}P-labeled 3'-phosphorothioate-$(dT)_3$. Lanes 0 – 2000: 3'-phosphorothioate-$(dT)_3$, 5'-iodo-$(dT)_4$, template strand $(dA)_7$, and proflavine at a concentration corresponding to the number above the lanes, in units of μM. All reaction mixtures were incubated for 1 h at 22ºC and loaded in duplicate. Marker lane where $(dT)_7$ strand is the most intense band with the lowest migration.

5.3.2. Linear Product Formation over the Duration of Ligation Experiments

A decline in reaction rate with increasing proflavine concentrations could be due to several factors such as inhibition of coupling chemistry, destabilization of ligation complex assembly and aggregation/precipitation of reactants by proflavine. Experiments were carried out to test if there was a decline in ligation activity with time due to limiting amounts of reaction components over time. Ligation reactions shown in Figure 5.2 had a 1 hour incubation period at 22ºC. We have determined reaction rates for different proflavine concentrations over the duration of 1 hour time period. A representative gel image in Figure 5.3A shows product formation at different time intervals from 0 to 60 minutes for a 100 μM bp ligation complex. Figure 5.3B shows rate profiles over a 1 hour time period for each proflavine concentration.

Observation of linear rate profiles is indicative of the fact that reactions are not limited by substrate or template strands. At concentrations above 1 mM proflavine, we observed sample clouding which is suggestive of an aggregation or precipitation phenomenon. It is entirely possible that nucleic acid substrates or template strands precipitate out of solution due to high concentration of proflavine. Loss in linear product formation is observed for a 1250 μM proflavine sample after 30 minutes. It is possible that as a function of time, precipitation of reaction components also increases in magnitude.

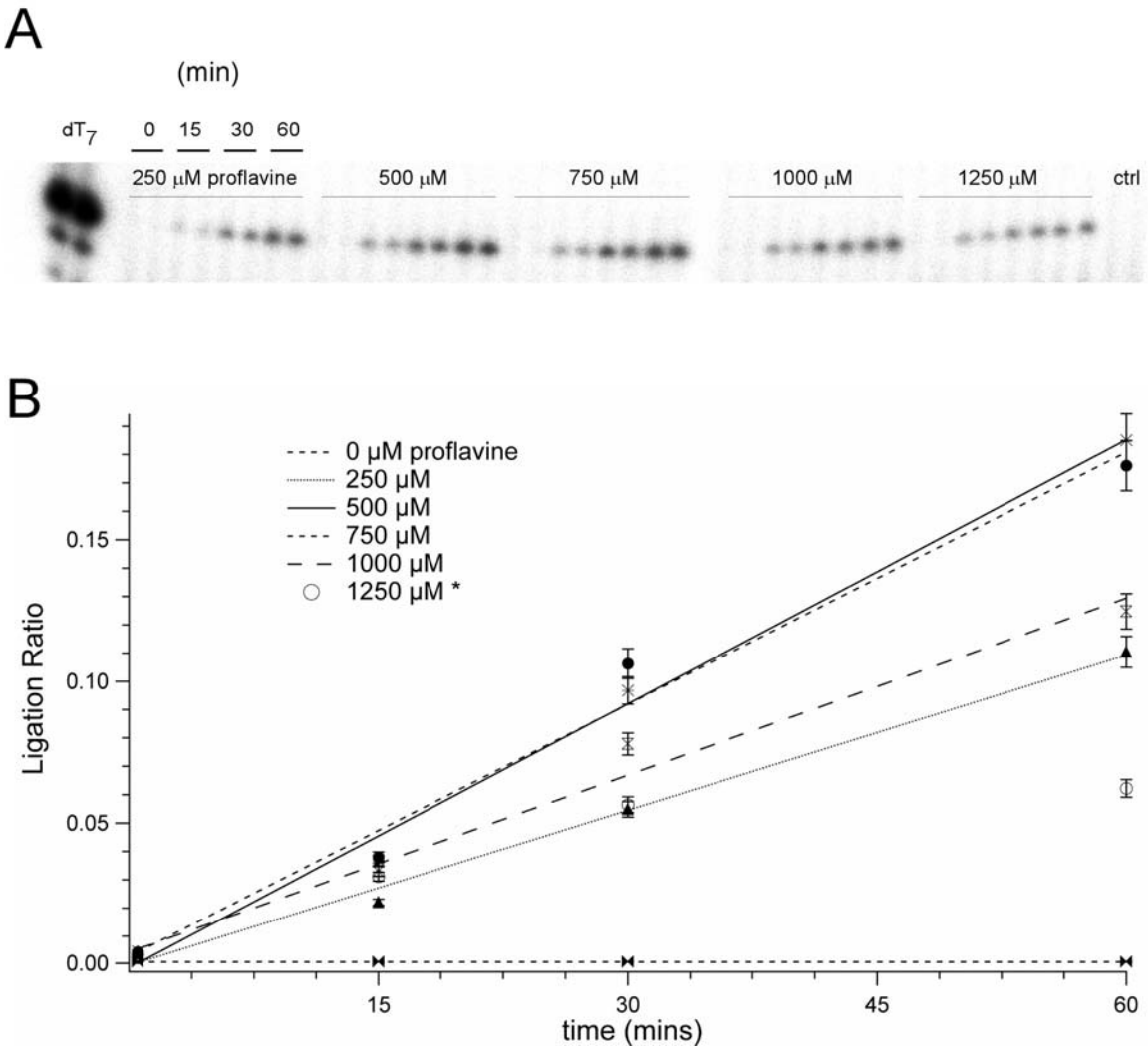


Figure 5.3. Kinetic studies of a proflavine-mediated ligation system. (A) Denaturing PAGE analysis illustrating product formation at different concentrations of proflavine at time points over the course of 1 hour. All samples were loaded in duplicate. Experimental conditions: 100 μM bp ligation complex, 10 mM Tris (pH 8.0), 100 mM 2-mercaptoethanol at 22ºC, proflavine concentration indicated above the lanes. (B) Quantitative analysis of the kinetic study. Ligation ratio was plotted as a function of time. Quantified gel bands were divided by a lower mobility (non ligation active) band in the gel to get the ligation ratio. The lower mobility band (Figure 5.2) stays constant for all lanes in which equal amount of radiolabeled substrate is added and serves as an internal standard for our reactions. Data were fit to a line function and display linearity over the course of 1 hour, except for reactions containing 1250 μM proflavine.

5.3.3. Circular Dichroism Spectroscopy Illustrates Proflavine-Mediated Assembly of a Ligation Active Complex

We have used circular dichroism spectroscopy to illustrate that proflavine interacts with substrate and template strands and assembles a ligation active complex. Proflavine is achiral and therefore does not by itself produce a circular dichroism signal. However, upon interaction with the chiral environment of nucleic acids, an induced CD is observed in the regions of proflavine absorption (e.g. ~ 450 nm). Figure 5.4A exhibits the CD spectrum of a model ligation complex (i.e., dA_7 + dT_4+ pdT_3p) with proflavine that is distinct in its features from a sample in the absence of proflavine. The appearance of induced CD bands in the region of proflavine absorption (~ 450 nm) and changes in the spectra at ~ 275 nm show that proflavine is interacting in a specific manner with the ligation active complex. Presence of positive induced CD bands has been previously shown to be an indicator of small molecule binding in the chiral environment of nucleic acids [86-88]. To determine if proflavine wsa interacting with template strands (dA_7) and model substrate strands (dT_4, pdT_3p) in a single complex, the spectra of individual components of the model ligation active complex were acquired in the presence of proflavine. That is, we have taken individual strands of the ligation complex (dA_7, dT_4, pdT_3p) and added proflavine to the sample (Figure 5.4B-D). Negligible changes are observed in the spectra in the presence or absence of proflavine.

Furthermore, the individual spectrum of the three single strands (dA_7, dT_4, pdT_3p) with proflavine were numerically added and compared to the ligation complex spectrum in the presence of proflavine. The summation spectra was clearly different from the model ligation complex (Figure 5.5A).

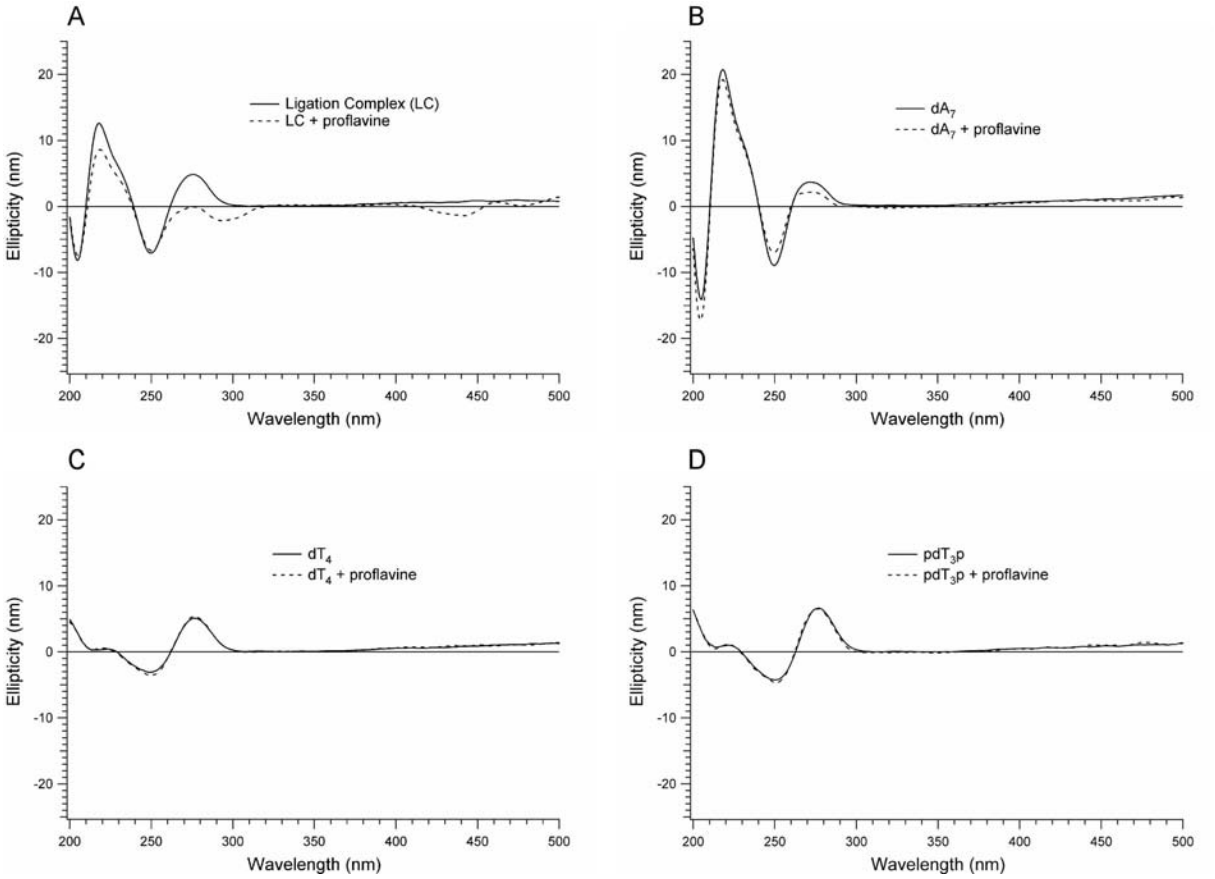


Figure 5.4. Circular dichroism spectra illustrating the formation of a model ligation active complex. (A) Ligation complex (LC) with (dashed) and without proflavine (solid) shows distinct changes in the spectrum from 260 nm to 310 nm and induced CD bands near 450 nm. (B-D) Controls (dA_7 , dT_4 , pdT_3p) with (dashed) and without (solid) proflavine do not show any appreciable changes in CD spectra illustrating negligible interaction and chiral ordering of proflavine. Minor changes in spectra are observed for the dA_7 sample in the presence of proflavine. It is possible that proflavine weakly interacts with a dA_7 strand because induced bands are not observed Figure 5.4B. Concentrations were 240 μM in base for each strand, 120 μM in proflavine, 10 mM Tris (pH 8.0). Spectra are an average of 3 scans acquired at 4ºC.

We have also measured CD spectra for the two samples dA_7 + dT_4 and dA_7 + pdT_3p in the presence of proflavine. Upon numerical addition of the two spectra and division by 2, we find that this summation spectrum bears features that resemble the ligation active complex spectrum (Figure 5.5B) but the spectrum of the model ligation complex is not reproduced perfectly by the summation of individual parts. It is not surprising to notice that this spectrum would have similarities to the model ligation complex because duplex formation is occurring in each sample despite the absence of one of the substrate strands.

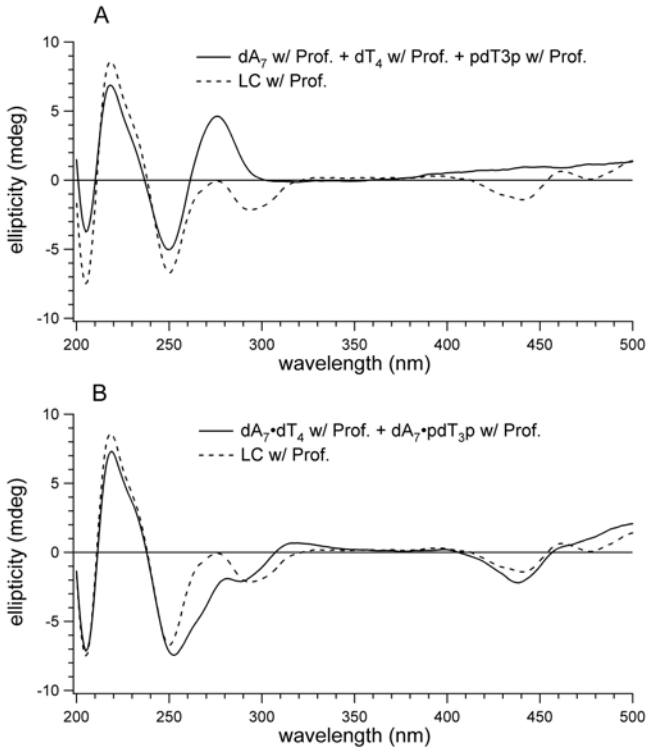


Figure 5.5. Summation CD spectra illustrating model ligation complex assembly in the presence of proflavine. (A) Spectrum of ligation complex with proflavine (solid) is distinct from a summation spectrum of the individual components of LC with proflavine. Absent are the changes near the 260 nm region and the induced bands near 460 nm. (B) Spectrum of individual substrates with template in the presence of proflavine were added and divided by 2 to yield the summation spectra (dashed) which shows some similarities to the ligation complex spectrum but is not identical. Concentrations were 240 μM in base for each strand, 120 μM in proflavine, 10 mM Tris (pH 8.0). Spectra are an average of 3 scans acquired at 4ºC.

5.3.4. Melting Temperatures Analysis and Ligation Complex Stability

The remarkable increase in product formation at 22ºC upon proflavine addition suggested to us that proflavine must be assembling a ligation complex where substrates align alongside template strands, thereby allowing chemical coupling between reactive groups to take place (Figure 3.2A). In addition, this ligation complex must be present at some equilibrium amount at room temperature (reaction temperature). Circular dichroism (CD) melting profiles show that the addition of proflavine up to a concentration allowed by NNEP (120 μM) leads to the assembly of a complex from the model (pdT_3p) and $(dT)_4$ substrates and $(dA)_7$ template with a melting transition at 22ºC (Figure 5.6). CD spectra at 4 and 22ºC are shown in Figure 5.7 for the model ligation complex in the presence and absence of proflavine. We have monitored change in CD at 275 nm for the T_m analyses which is the region of proflavine and DNA absorption. Proflavine induced CD bands (400 – 500 nm) were small in magnitude and therefore could not be used for melting analyses.

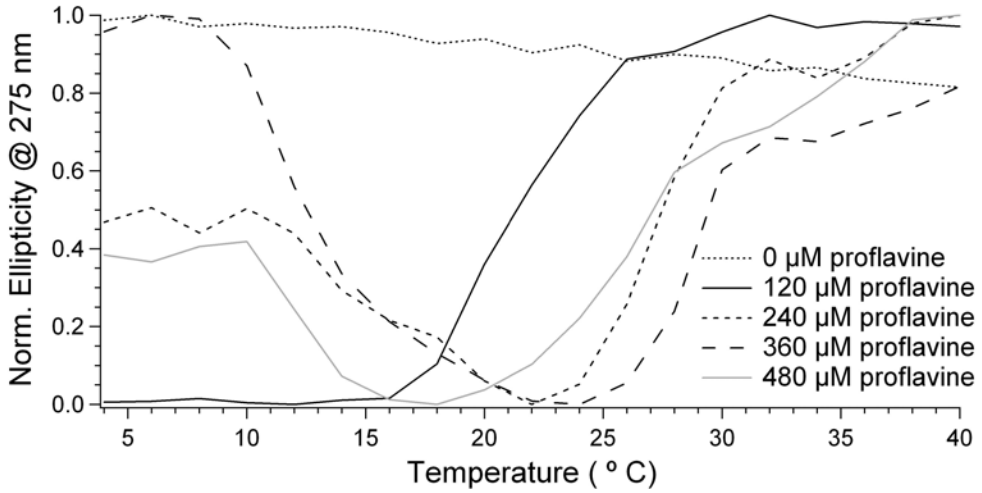


Figure 5.6. Circular dichroism (CD) melting profiles for a model of the ligation complex (dA_7 + dT_4 + pdT_3p) at 275 nm. Ligation complex concentration was 240 μM in bp. Proflavine concentrations are indicated in the legend.

Our gel ligation data suggests that the optimum assembly of a ligation complex occurs at a proflavine concentration beyond that allowed by the NNEP. We have also investigated the effect of higher proflavine concentrations on the stability of the complex. Addition of proflavine up to a concentration of 240 μM (2x NNEP) leads to an appreciable increase in the thermal stability of the complex (T_m = 27 ° C) (Figure 5.6). Further addition of proflavine (up to 360 μM) leads to a small increase in complex stability, which is followed by destabilization of the ligation complex at a concentration of 480 μM proflavine (Figure 5.6, Table 5.1). The significance of complex destabilization at higher proflavine concentrations is also consistent with our gel data which shows a decline in product formation at higher concentrations of proflavine (Figure 5.2).

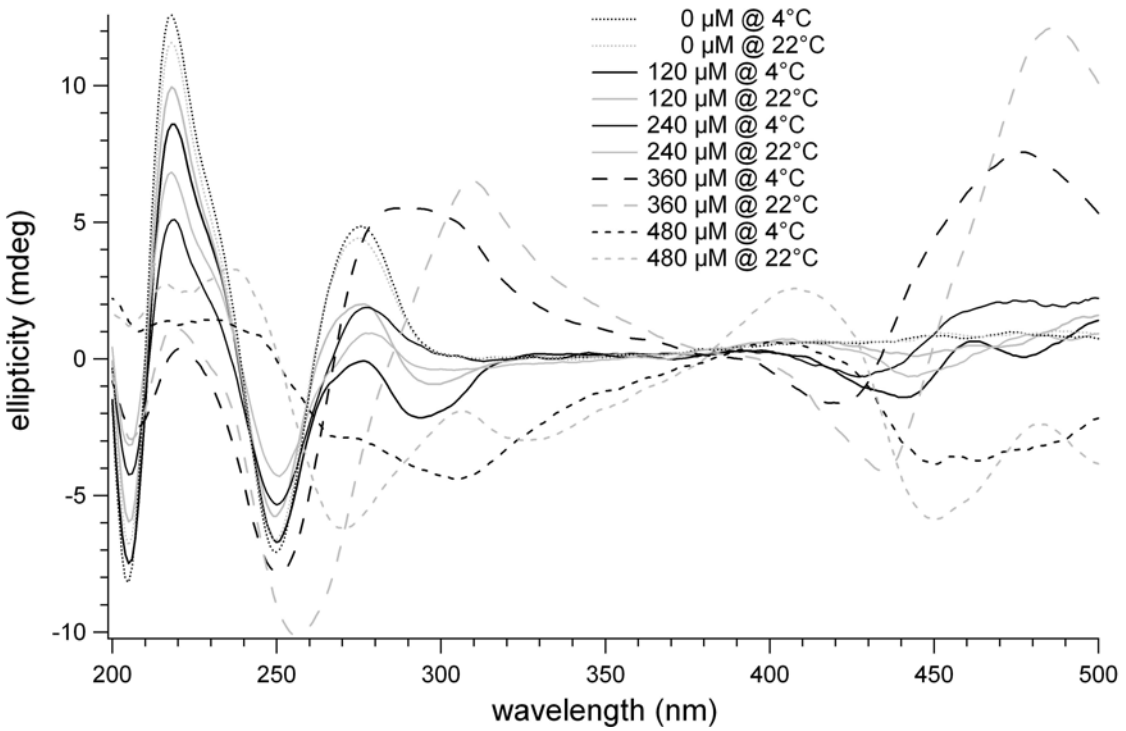


Figure 5.7. CD spectra of the model ligation complex in the presence of proflavine. Concentration of the model ligation complex was 240 μM in bp. Samples contained 100 mM Tris, pH 8.0.Spectra were collected at 4 and 22ºC and are an average of 5 scans.

CD melting profiles in Figure 5.6 depict an additional transition centered at 12ºC besides the ligation complex melting transition. Note that this transition is only apparent at concentrations beyond that allowed by NNEP for proflavine binding to the duplex. Furthermore, this transition increases in intensity in a concentration dependent manner up to 360 μM proflavine, followed by a decline at 480 μM proflavine. This observation suggests that an additional mode of proflavine binding to the ligation complex may also be prevalent.

Table 5.1: T_m Analysis of the Model Ligation Complex

Sample	*[Proflavine] μM*	*T_m (ºC)*
$(dA)_7.(dT)_4.p(dT)_3p$ (LC)	0	N/A
LC	120	22
LC	240	27
LC	360	29
LC	480	26
$(dA)_7.(dT)_4$	120	20
$(dA)_7.p(dT)_3p$	120	17
$(dA)_7.(dT)_3$	120	15

Table 5.1. Concentration of the 'model ligation complex' (LC) was 240 μM in bp. Proflavine concentration is indicated in the table for each melting experiment. T_m is melting of the ligation active duplex structure. Control melts (dA_7•dT_4), (dA_7•pdT_3p), and (dA_7•dT_3) contained 240 μM in base of template and substrate. These should be directly compared to LC with 120 μM proflavine.

5.3.5. Decline in Ligation Activity Suggests Additional Binding Modes of Proflavine

With the oligonucleotide $(dA)_7$ as a template, the proflavine-catalyzed ligation of the $(dT)_3$ and $(dT)_4$ substrates showed a substantial decline in ligation at higher concentrations of proflavine which suggests that a non-active ligation complex becomes favorable. The nature of this putative non-active ligation complex is not self-evident. A decline in $(dT)_7$ product formation could be due to 1) inhibition of the ligation chemistry at higher concentrations of proflavine and 2) disruption of ligation complex assembly. It has been demonstrated that in addition to intercalation, proflavine binds to nucleic acids by a weaker binding process which involves outside stacking of proflavine molecules with each other alongside the sugar-phosphate backbone of nucleic acids [90, 160]. We wanted to investigate whether proflavine interacts with the ligation active complex via additional modes of binding besides intercalation which may lead to a decline in ligation activity. For this work, we have undertaken rigorous CD titration studies of proflavine with a stable $(dA)_{16}$•$(dT)_{16}$ duplex and an unstable $(dA)_7$•$(dT)_7$ structure. Figure 5.8 shows the CD spectra of a titration study involving proflavine and $(dA)_{16}$•$(dT)_{16}$ duplex. Change in the CD spectra of a DNA sample can indicate a change in the binding behavior of proflavine to the nucleic acids. We have plotted CD signal intensity at specific wavelengths as a function of R (ratio of the number of nucleosides per proflavine molecules in sample solution) to demonstrate the stoichiometry at which a change in binding behavior occurs. Titrations were carried out in two distinct ways: 1) DNA was titrated into a constant proflavine stock where the R value increases as a function of increasing DNA concentration 2) Proflavine was titrated into a constant DNA stock where R decreases as a function of increasing proflavine concentration.

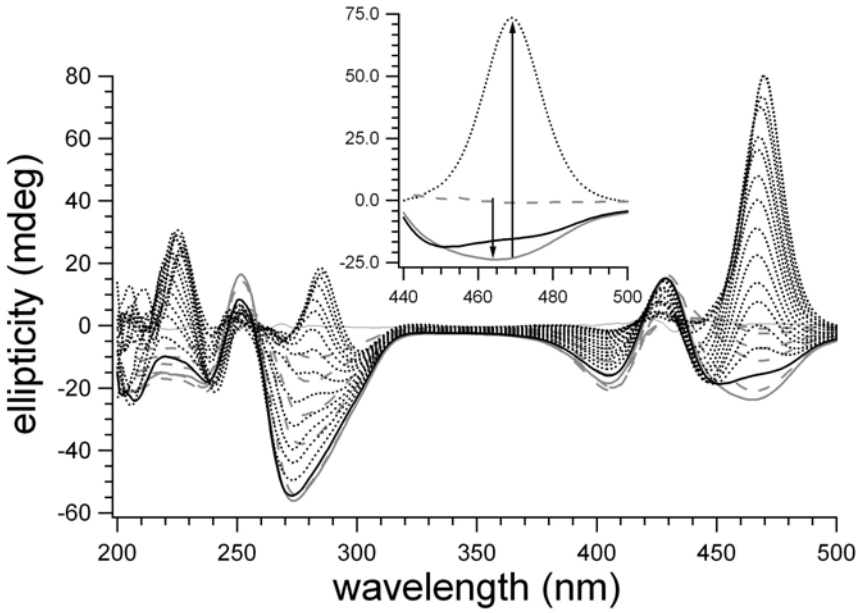


Figure 5.8. CD spectra from the titration of a $(dA)_{16}$•$(dT)_{16}$ duplex sample into a proflavine stock. Sample concentration was 250 μM proflavine, 10 mM Tris (pH 8.0). Titrant solution concentration was 10 mM bp $(dA)_{16}$•$(dT)_{16}$. Sample was maintained at 4ºC during the duration of the titration. Dashed gray spectra approaching the solid gray spectrum are the first additions of DNA into a proflavine stock. We observe a decrease in intensity at 465 nm and the local minima is represented by the solid gray spectrum. Further increase in DNA concentration leads to an abrupt reversal in the direction of CD spectra. Solid black line represents the first increase in intensity at 465 nm and further increments are illustrated by the dotted black spectrum approaching a local maximum at 465 nm. Inset graph clearly illustrates the changes in CD spectra over the course of the titration.

Titration of a $(dA)_{16}$•$(dT)_{16}$ duplex into a 250 μM proflavine sample shows a sudden change in CD spectra at R = 1 (# of nucleosides = # of proflavine) which is represented by the solid gray spectrum in Figure 5.8. The inflection point observed in the graph in Figure 5.9A is suggestive of distinct modes of proflavine binding below and above R = 1. We also observe subtle changes in the direction of the CD spectra at R = 2, 3, and 4 which could indicate binding of 1 proflavine per 2, 3, and 4 phosphate atoms, respectively. We have conducted this type of titration analysis at several different proflavine concentrations and observe very similar results (Figure 5.9A). Proflavine was

also titrated into a 100 µM bp $(dA)_{16}$•$(dT)_{16}$ duplex at 22ºC (Figure 5.9B). Duplex $(dA)_{16}$•$(dT)_{16}$ is stable up to 24ºC even in the absence of proflavine. Therefore, a titration where either proflavine is added to a DNA stock or vice versa should yield similar results. We observe a sharp inflection at a value of R = 1 which is consistent with the idea that there are additional modes of proflavine binding besides intercalation.

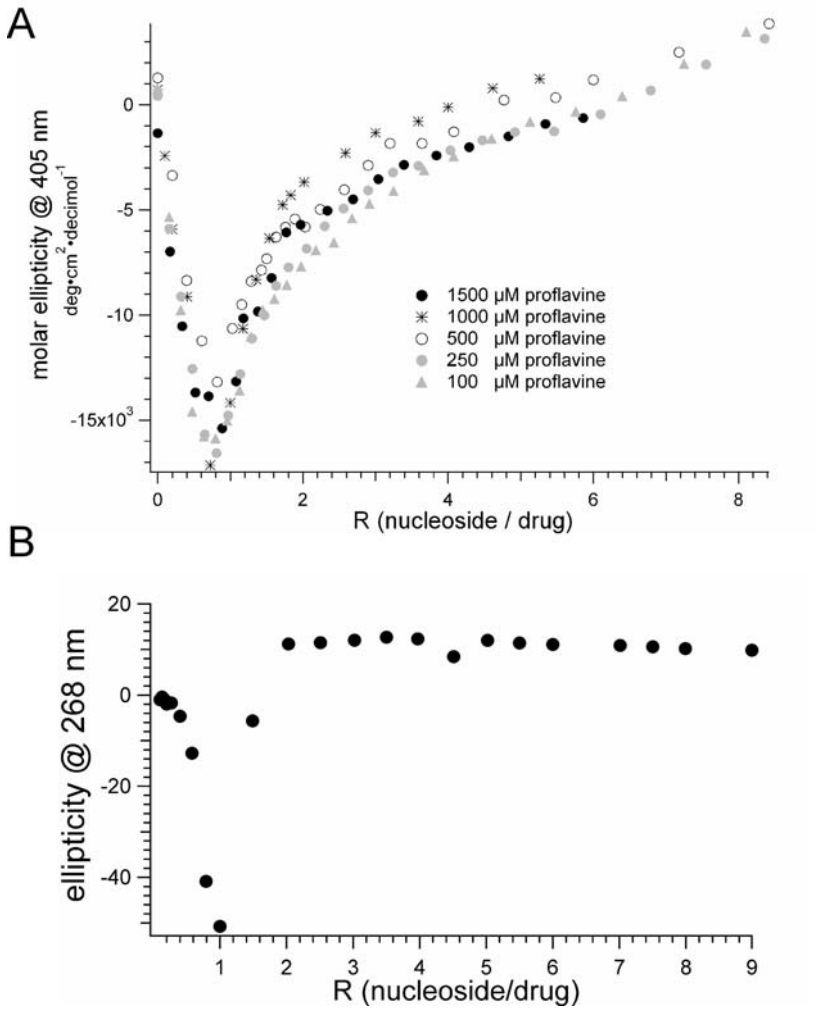


Figure 5.9. CD titration analysis of proflavine and $(dA)_{16}$•$(dT)_{16}$ duplex. (A) Molar ellipticity at 405 nm was plotted as a function of R (nucleoside/drug) from several experiments at 4ºC where a 10 mM $(dA)_{16}$•$(dT)_{16}$ duplex was titrated into various proflavine samples. (B) Ellipticity at 268 nm was plotted as a function of R. Data was obtained from a CD titration at 22ºC where a 5 mM proflavine sample was titrated into a 100 µM bp $(dA)_{16}$•$(dT)_{16}$ duplex.

We have carried out similar analyses for a $(dA)_7$•$(dT)_7$ duplex which serves as an excellent model system for comparison to the ligation complex because proflavine is requisite to the assembly of a stable duplex in both systems. For a $(dA)_7$•$(dT)_7$ system, titrations were carried out where proflavine is added to a DNA stock. In this type of a titration, R value decreases as a function of increasing proflavine concentration. It was necessary to conduct the titration in this manner because addition of a small quantity of proflavine to $(dA)_7 + (dT)_7$ assembles the duplex and subsequent proflavine upon addition binds to the assembled structure. On the other hand, addition of a small quantity of DNA to a large concentration of proflavine leads to interaction of single strands with proflavine that is apparently more stable than the duplex structure. The single strand interactions with proflavine may also represent a kinetic trap rather than a thermostable state.

Figure 5.10 shows $(dA)_7 + (dT)_7$ titration with proflavine at 4ºC and 22ºC. A sharp inflection point is observed at R = 1. A change in the direction of CD spectra beyond R =1 illustrates a change in the binding mode of proflavine to a $(dA)_7$•$(dT)_7$ duplex similar to the results obtained for a $(dA)_{16}$•$(dT)_{16}$ duplex.

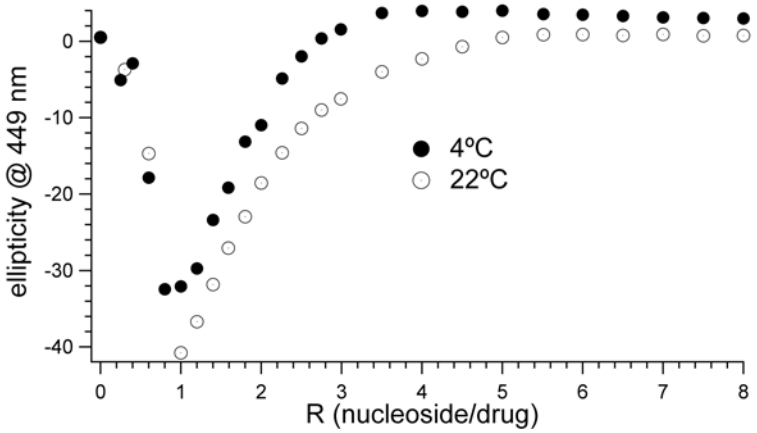


Figure 5.10. CD titration analysis of a $(dA)_7$•$(dT)_7$ duplex with proflavine. A 5 mM proflavine stock was titrated into a 100 μM bp $(dA)_7$•$(dT)_7$ duplex at 4 and 22ºC. CD spectra were obtained and ellipticity at 449 nm was plotted as a function of R.

Titrations of proflavine with the model ligation complex have been difficult to study using circular dichroism spectroscopy. Although the ligation system is similar to a dA)$_7$•(dT)$_7$ system in its base pairing potential, it is different in that there is a break in the backbone that is expected to greatly reduce stability.

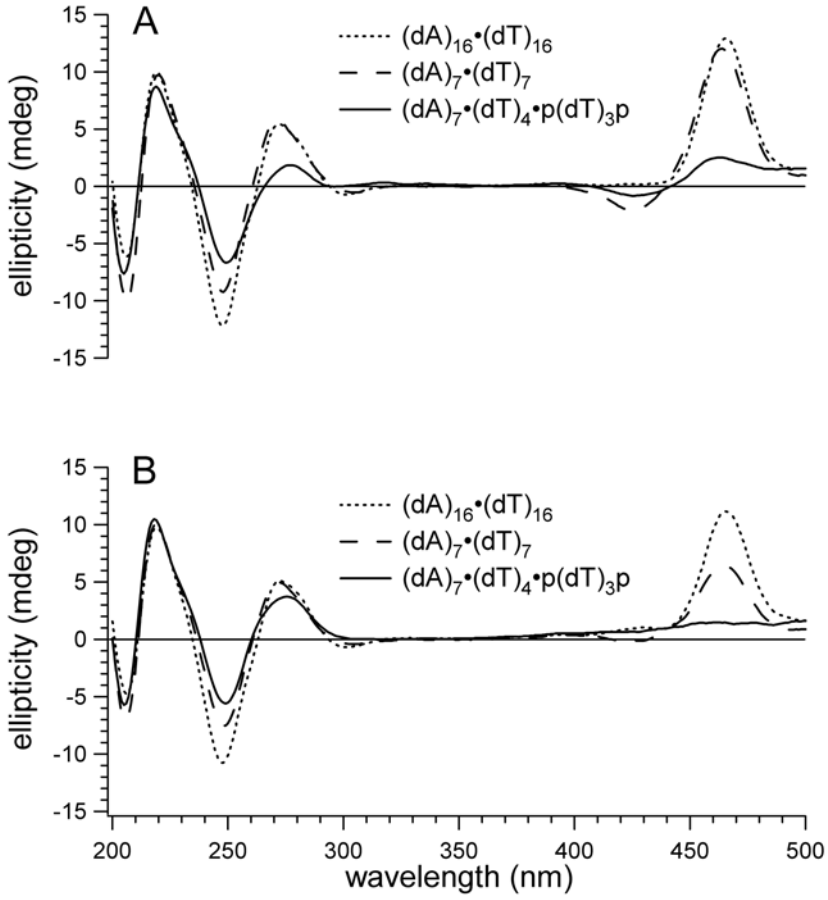


Figure 5.11. CD spectra illustrate proflavine-mediated assembly with nucleic acid systems of varying stability. (A) At 4ºC, (dA)$_{16}$•(dT)$_{16}$ and (dA)$_7$•(dT)$_7$ structures in the presence of proflavine form an assembled structure equal in magnitude to each other (note CD intensity at 465 nm). Assembly of a ligation complex (solid black) is diminished significantly in comparison to the 16-mer and 7-mer duplex systems. (B) CD spectra at 22ºC illustrate that the order of stability of the different systems can be determined by observing the magnitude of the induced CD band at 465 nm. Solution conditions were 240 μM bp, 120 μM proflavine, 10 mM Tris (pH 8.0).

The induced CD bands in the proflavine absorption region for the ligation active complex are a fraction in intensity of that observed for a $(dA)_{16}$•$(dT)_{16}$ duplex or a $dA)_7$•$(dT)_7$ structure (Figure 5.11A).The lack of induced bands is even more dramatic in a comparison between the three nucleic acid systems at 22ºC (Figure 5.11B). From CD spectra, it can be inferred that the magnitude of assembled structure follows the following order: $(dA)_{16}$•$(dT)_{16}$ > $dA)_7$•$(dT)_7$ >> $(dA)_7$•$(dT)_4$•$(dT)_3$.

5.3.6. Proflavine Binding Is Coupled to Duplex Stability

Our data has shown that proflavine-mediated assembly of a ligation active complex has thermal stability in excess of 22ºC. We wanted to determine the binding constants of proflavine to nucleic acids under the solution conditions of our study. Addition of proflavine up to that allowed by NNEP (1 proflavine per 4 bases) yields a strong, positive induced CD band with local maxima at 465 nm. Release of proflavine from the duplex structure should be accompanied by a decrease in the CD band at 465 nm. For the determination of proflavine binding constant, we have undertaken dilution studies where we have added 100 μM proflavine to a 400 μM base $(dA)_{16}$•$(dT)_{16}$ duplex. Dilution of a sample leads to a subsequent decrease in signal intensity. However, our experimental setup allows for a 100-fold dilution of samples without compensating signal strength by simply increasing the path length of the CD cuvettes. As long as proflavine is tightly bound to the duplex, subsequent dilutions do not change CD signal intensity on a per molar basis. A decrease is signal intensity (beyond that corresponding to the dilution factor) is indicative of a transition during which bound proflavine begins to dissociate from the duplex structure. The positive CD band at 465 nm in Figure 5.12A decreases in

intensity as a function of increasing dilution of sample and the induced CD band disappears when proflavine is fully released from the duplex sample. Inset graph in Figure 5.12A shows normalized molar ellipticity at 465 nm plotted as a function of proflavine concentration. This analysis shows that proflavine binds to a stable structure with a binding constant of ca. 5 μM. Thermodynamic predictions as well as our data shows that the $(dA)_{16}$•$(dT)_{16}$ duplex structure in the absence of proflavine has a stability of approximately 24ºC even upon a 100-fold dilution (4 μM base). [161]. Therefore, the binding constants that have been determined here are due to the release of proflavine from a stable duplex, and not coupled to a transition involving the melting of the duplex structure.

It has been reported that small molecules can also bind to nucleic acid structures with a weaker binding process that involves interactions with the phosphate groups of the nucleic acid backbone [69, 70, 90]. This binding process also seems to be playing a prominent role in the assembly of a ligation active complex. Titration studies show a sharp inflection at a binding ratio of 1 proflavine molecule per nucleoside (Fig 5.9). We wanted to determine the binding constants for this additional mode of binding observed here. Figure 5.12B shows that upon addition of a 400 μM proflavine to a 400 μM per base $(dA)_{16}$•$(dT)_{16}$ duplex, a broad negative induced signal is observed at ~ 470 nm. This peak is significantly different than the positive peak observed for a 4:1 (proflavine:nucleoside) complex. Nevertheless, dilution studies can be carried out in a similar manner and we find that the binding constant for the release of proflavine from a $(dA)_{16}$•$(dT)_{16}$ sample is ~ 15 μM (inset Figure 5.12B).

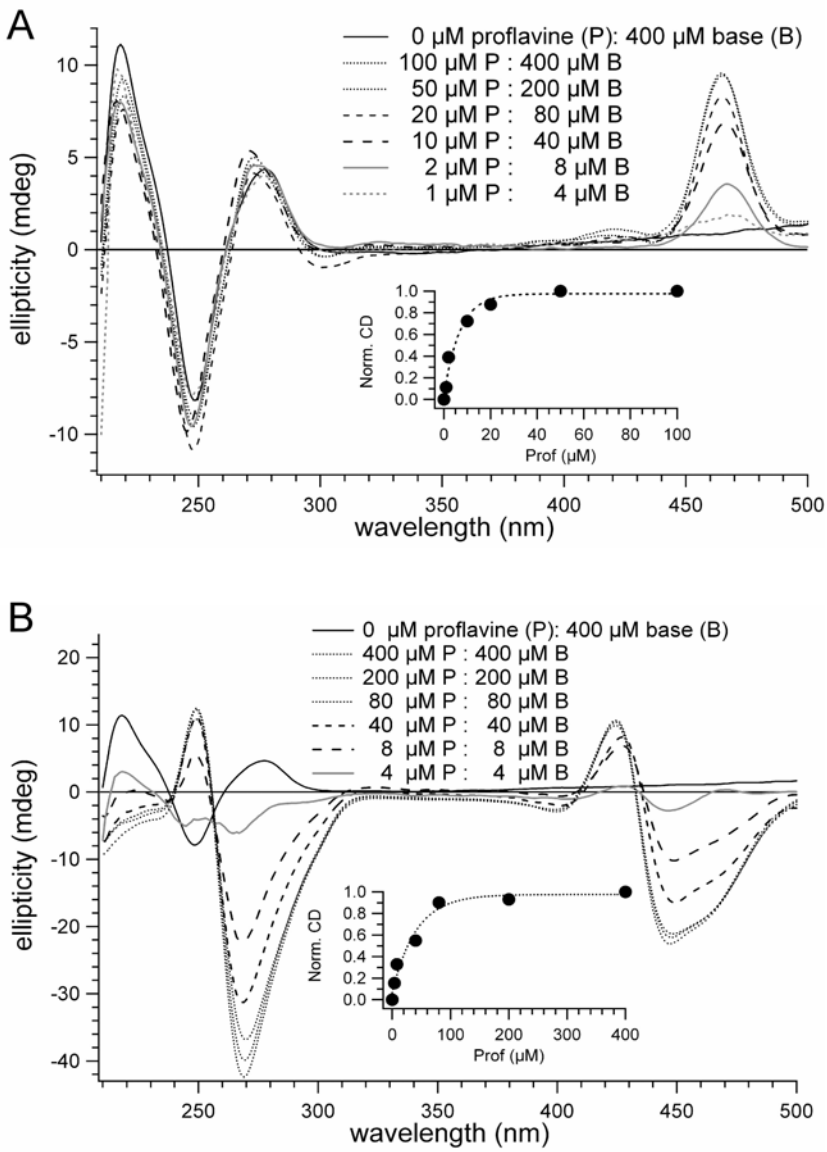


Figure 5.12. CD spectroscopy for the determination of K_d for proflavine binding to $(dA)_{16}$•$(dT)_{16}$ duplex. (A) Dilution studies were carried out on a sample with initial concentration of 400 µM base, 100 µM proflavine (NNEP, R = 4), 10 mM Tris (pH 8.0). 1 mm path length cell was used for the starting sample. Subsequent samples with dilutions of 1:2, 1:5, 1:10, 1:50, 1:100 were measureD in 2 mm, 5 mm, 10 mm, 50 mm, 100 mm cells, respectively. (inset to A) Normalized CD at 465 nm was plotted as a function of proflavine concentration to determine binding constant. (B) Dilution studies were carried out on a sample with initial concentration of 400 µM base, 400 µM proflavine, 10 mM Tris (pH 8.0). This sample represents a model complex with R =1 where nucleosides equal the number of proflavine molecules in the sample. Dilution and sample measurements were carried out in an identical manner to A. (inset to B) Normalized CD signal at 447 nm was plotted as a function of proflavine concentration for binding constant determination. CD signal in the inset plots represents the molar ellipticity at the indicated wavelengths.

We have conducted similar dilution studies with a sample containing $(dA)_7$ and $(dT)_7$ strands. This system closely mimics our ligation active complex $[(dA)_7•(dT)_4•(dT)_3]$ because single strands $(dA)_7$ and $(dT)_7$ form a stable $(dA)_7•(dT)_7$ duplex structure only in the presence of proflavine. We have taken 1000 μM base $(dA)_7•(dT)_7$ and added 250 μM proflavine. The binding constant determined for a proflavine-bound $(dA)_7•(dT)_7$ duplex is ~ 50 μM proflavine where the DNA concentration is 200 μM base . Note that the K_d determined for a $(dA)_7•(dT)_7$ system is for the melting of the entire complex (proflavine-bound duplex) because release of proflavine also leads to the melting of the $(dA)_7•(dT)_7$ duplex into single strands. According to thermodynamic predictions, the strand dissociation constant (K_d) for a $(dA)_7•(dT)_7$ duplex is ~ 80 mM in base [161]. Observation of a stable duplex structure at 200 μM base upon addition of proflavine suggests that the "effective concentration" of the duplex structure can be shifted ~ 400-fold by addition of proflavine.

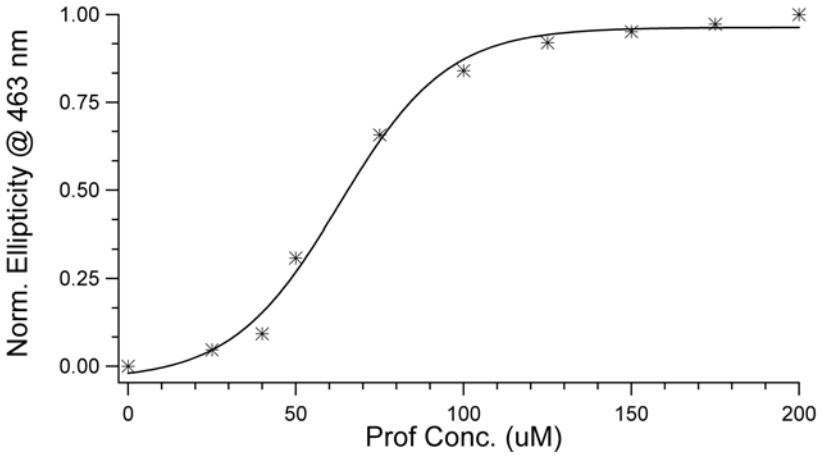


Figure 5.13. Determination of K_d for a proflavine $(dA)_7•(dT)_7$ complex. Normalized molar ellipticity at 463 nm is plotted as a function of proflavine concentration. Dilution studies at room temperature were carried out on a sample containing 200 μM proflavine and 400 μM bp $(dA)_7•(dT)_7$, 10 mM Tris, pH 8.0.

Binding constant determination for the $(dA)_7$•$(dT)_7$ duplex with proflavine concentrations beyond that allowed by nearest neighbor exclusion principle has been difficult to study due to the single strand interactions with proflavine. Visible aggregation and sample precipitation is apparent in a $(dA)_7$ + $(dT)_7$ sample at $R \leq 2$ (number of nucleoside per proflavine molecules).

5.3.7. Ligation Profiles at 22ºC for Varying Concentrations of Ligation Complex

Ligations as a function of proflavine concentration were undertaken for several different concentrations of the ligation complex. The purpose of these experiments was to determine how ligation activity changes as a function of increasing concentration of ligation components. All samples contained an equal amount of the radiolabeled dT_3 substrate. Experiments where the radiolabeled amount is kept constant are a direct reporter of enhancement or suppression of ligation activity upon changes in reaction parameters. We have investigated ligation substrate and template concentrations of 1 μM to 1000 μM bp. The resulting data is presented in Figure 5.14.

Dotted line in Figure 5.14 is the ligation profile for a 1 μM bp complex. We observe an increase in ligation which tends to plateau at higher concentrations of proflavine. A 10-fold increase in the concentration of ligation complex (dashed) leads to a significant increase in ligation activity which is followed by a decline. It appears that the 1 μM bp complex is limited by DNA strand concentration because increasing the concentration of the DNA strands results in a significant increase in the product strand formation. A 100 μM bp complex (solid bold) shows the highest ligation activity. However, this activity is not 10-fold higher than the 10 μM bp complex (dashed) which

points towards a mechanism other than concentration of strands that is limiting the ligation activity. Further increase in ligation complex concentration leads to a decline in ligation activity. A 1000 μM bp complex shows ligation activity that is similar in magnitude to a 1 μM complex.

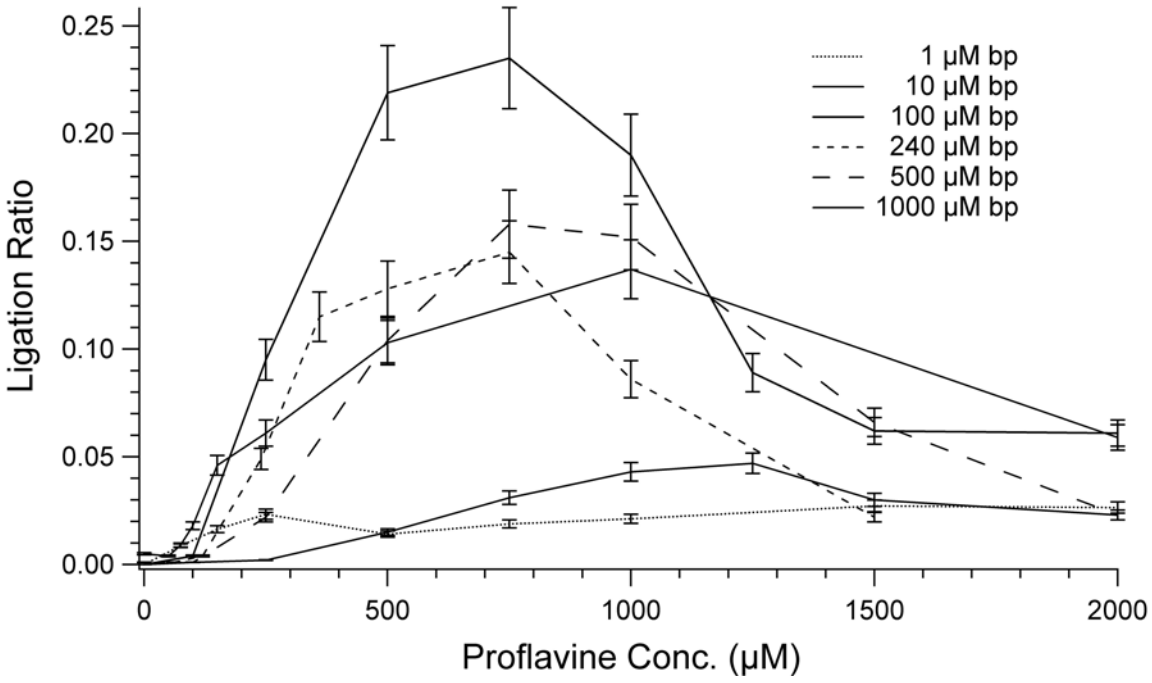


Figure 5.14. Ligation profiles as a function of proflavine concentrations. Experiments were conducted at 22ºC with incubation time of 1 hour. All samples contained 100 mM 2-mercaptoethanol, 10 mM Tris, and 0.1 μM base of radiolabeled phosphorothioate-dT_3 strand. DNA reaction component concentrations are indicated in the figure legend.

A closer look at the ligation profiles shows that the half maximum ligation for a 1 μM complex occurs at a higher proflavine concentration than the 10 μM bp and 100 μM bp complexes. We also note that the half maximum ligation point for 500 μM and 1000

μM bp profiles is again shifted to a higher concentration of in comparison to a 100 μM bp complex. These observations suggest that ligation activity is controlled by the binding constants of proflavine to the ligation active structure. We also visually observe opaque solutions for ligation complex concentrations greater than 500 μM in bp. This observation suggests that an aggregation/precipitation phenomenon is limiting ligaiton above 500 μM in DNA bp.

5.3.8. HPLC Purification and MS Analysis of the Modified Substrates

5.3.8.1. 3'-phosphorothioate, 5'-phosphate-$(dT)_3$

Figure 5.15 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time (min) for the purification of 3'-phosphorothioate, 5'-phosphate-$(dT)_3$ oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 5% acetonitrile, 50 min = 20% acetonitrile. Flow rate was 4.6 ml/min.

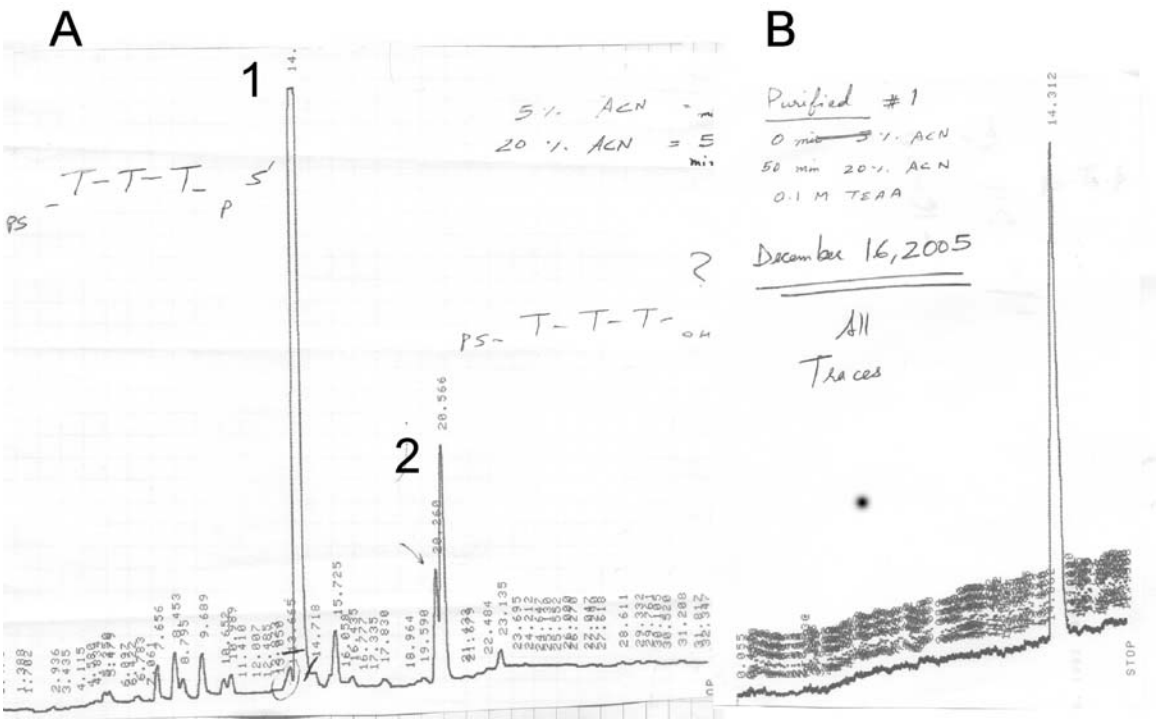


Figure 5.15. HPLC purification of 3'-phosphorothioate, 5'-phosphate-$(dT)_3$.
(A) Chromatograph of a crude sample is shown. Two fractions (#1 and #2) were collected. (B) Fraction # 1 (major) was run on the column a second time to confirm purity.

Mass spectrum of a purified 3'-phosphorothioate, 5'-phosphate-$(dT)_3$ sample is shown in Figure 5.16. Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 1025.6 g/mol. M-H peak is observed at 1024.6 g/mol. We also observe M - 2H (512.0 g/mol) and M - 3H (341.0 g/mol) peaks which indicate multiple charged species in the sample.

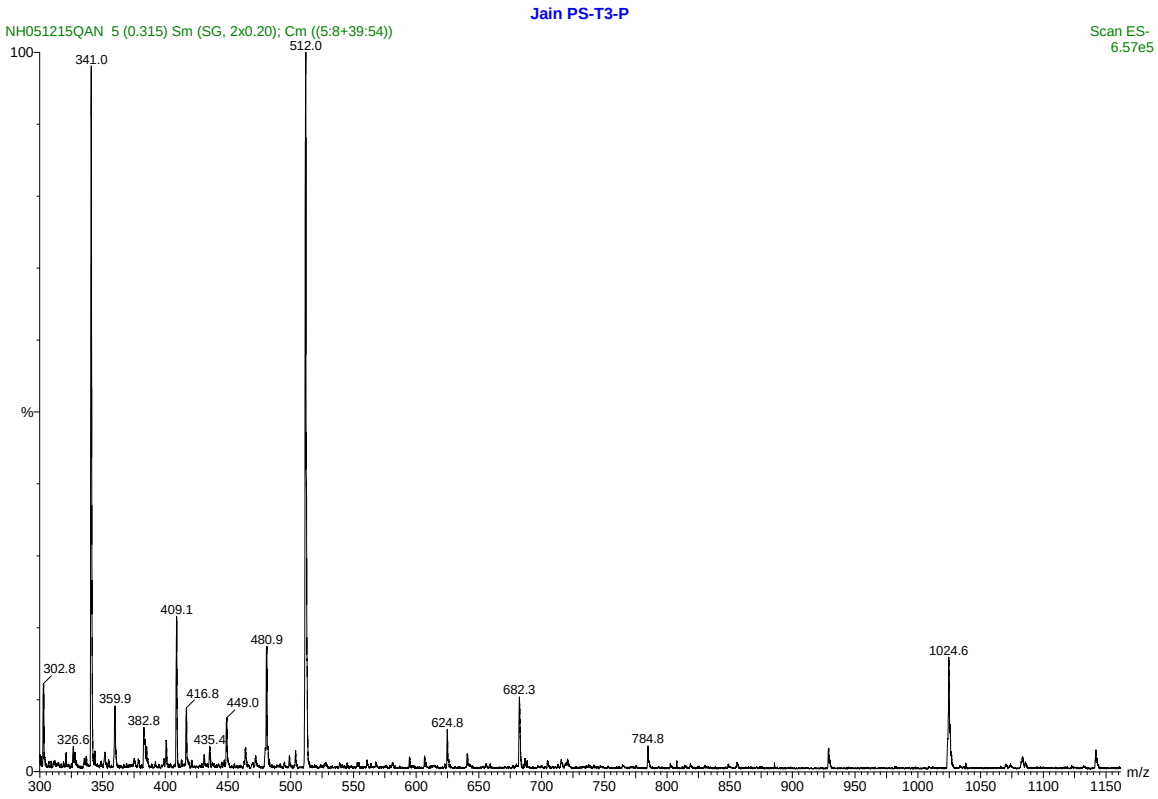


Figure 5.16. Mass Spectrum of a 3'-phosphorothioate, 5'-phosphate-$(dT)_3$ oligonucleotide.

5.3.8.2. 3'-phosphate, 5'-phosphate-$(dT)_3$

Figure 5.17 shows an HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time (min) for the purification of a 3'-phosphate, 5'-phosphate-$(dT)_3$ oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 µ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 5% acetonitrile, 50 min = 20% acetonitrile. Flow rate was 4.6 ml/min.

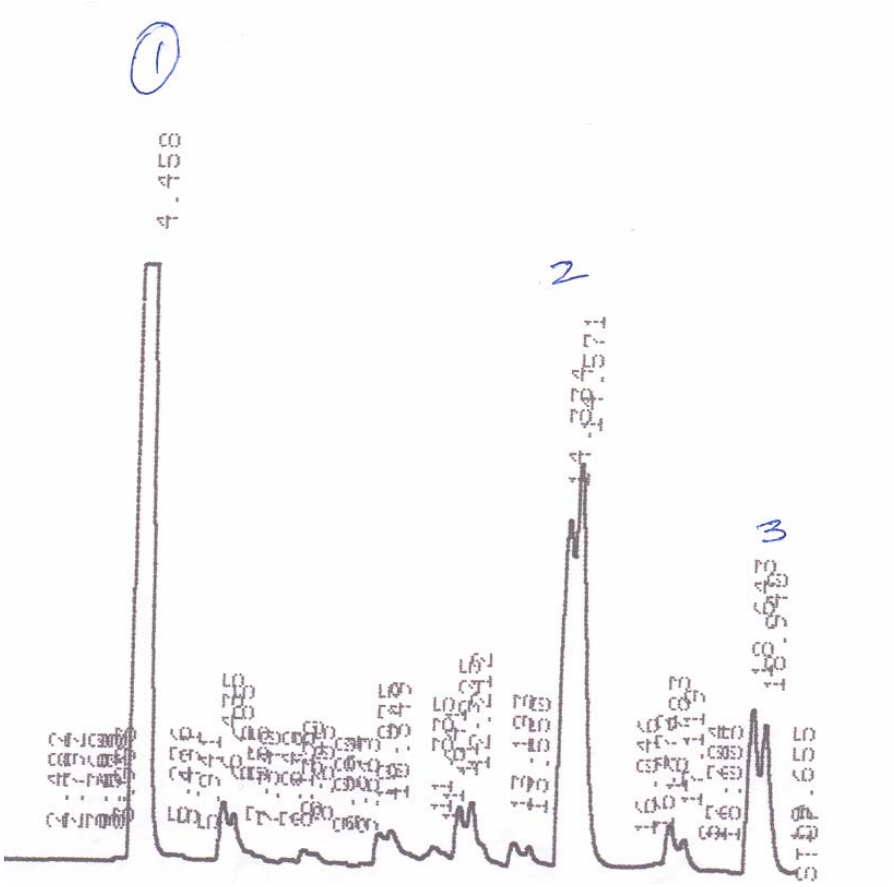


Figure 5.17. HPLC purification of a 3'-phosphate, 5'-phosphate-$(dT)_3$ oligonucleotide. Chromatograph of a crude sample is shown where three major fractions were collected. All fractions were submitted for MS analysis.

Mass spectrum of a purified 3'-phosphate, 5'-phosphate-$(dT)_3$ oligonucleotide sample is shown in Figure 5.18. Peak 2 from the HPLC purification was found to be the compound of interest even though it was not the major fraction. Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 1010.0 g/mol. Two major peaks are observed in the spectrum. M-H peak is observed at 1008.8 g/mol and M - 2H peak is noted at 504.0 g/mol. M – 2H peak indicates a doubly charged species.

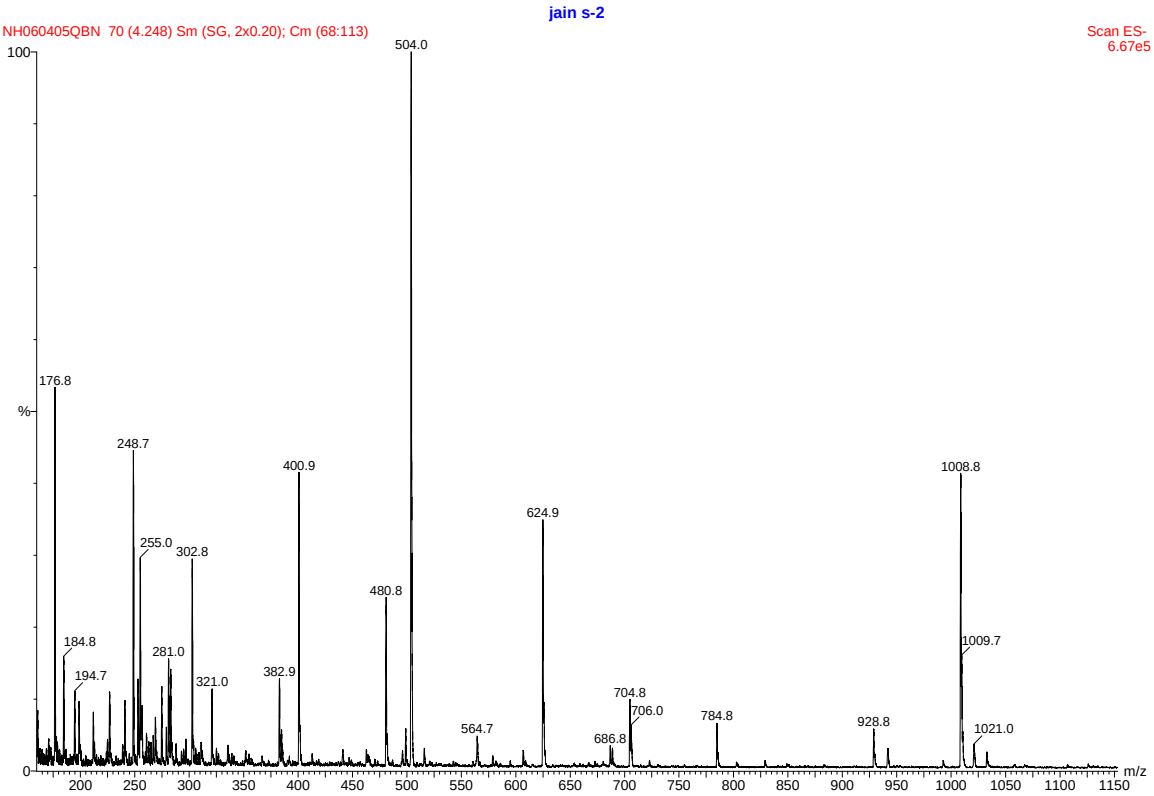


Figure 5.18. Mass spectrum of a 3'-phosphate, 5'-phosphate-$(dT)_3$ oligonucleotide.

5.4. CONCLUDING REMARKS

Data from several different experimental techniques demonstrate the remarkable utility of proflavine in the stable assembly of a ligation active ternary complex (A•T base pairs only) consisting of trimer and tetramer oligonucleotides pairing to a heptamer template. A strong interplay exists between coupling chemistry and ligation complex assembly. From our gel ligation experiments, product formation is detected which is a results of stable assembly as well as successfully coupling. Although circular dichroism spectroscopy reports only on the assembly, it has nevertheless proved to be a vital tool in understanding the details of the ligation system. In our previously published work, we found that additional molecules of proflavine bound to the ligation active complex which leads to a decline in ligation activity. Here, we have discovered that proflavine binds to nucleic acid structures by additional modes of binding. We have also been able to determine binding stoichiometry, binding affinities and melting temperatures of a model ligation complex. Our studies have revealed a complex binding behavior of proflavine with nucleic acids that involves an interplay between binding constants and the stoichiometry of proflavine binding to nucleic acids.

We have found that ligation activity can be limited by several factors. Proflavine must be present at a concentration which is at or above its binding constant to the ligation complex. Reactions must not be limited by the concentration of nucleic acid substrates and templates. In addition, proflavine must not bind to ligation complex at concentrations where aggregation occurs or where proflavine binding leads to the formation of ligation-inactive structures. Keeping in view the above findings, future experiments should be able to more fully exploit small molecule mediated assembly of nucleic acids.

CHAPTER 6

SMALL MOLECULE MEDIATED ASSEMBLY OF NUCLEIC ACIDS: NEW DEVELOPMENTS AND FUTURE DIRECTIONS

6.1. INTRODUCTION

Small molecule intercalation of nucleic acid base pairs represents a new and powerful method for enhancing the ligation rate of short oligonucleotides in a template-directed reaction [85]. Our studies with proflavine have demonstrated that besides intercalation, additional binding modes of proflavine to nucleic acids are necessary, in some cases, for the assembly of a ligation active complex. Our work with coralyne illustrates the remarkable ability of a small molecule in promoting novel secondary structures However, we have also found that assembly of nucleic acids using small molecules is significantly more complicated than we first anticipated [86-88].

Efforts to condense mononucleotides using numerous activation chemistries have been reported in the past few decades [35, 38, 42, 47, 162]. In particular, 2-methylimidazole phosphate (2-MeIMP) activation chemistry used by Orgel's group has been cited extensively in literature [35-37, 52]. However, these condensation reactions have been severely limited in their scope due to strict sequence requirements. In addition, experiments using these activation groups rely on the presence on high concentrations of Na^{+} and Mg^{++}. We have sought to determine whether small molecule binding can lead to the coupling of 2-MeImp monomers under solution conditions where these reactions have failed in the past.

Recent efforts in the area of small molecule binding to nucleic acids in our laboraotry have focused primarily on $(dA)_n$•$(dT)_n$ nucleic acid structures. Here, we show that assembly as well as ligation can be extended to nucleic acids G•C base pairs. We have also designed a stable ligation system which is not dependent upon small molecule binding for assembly and coupling of substrates. Experiments with a stable ligation system have revealed that proflavine, at high concentrations, can in fact be deleterious to ligation. We also find that coupling chemistry is significantly enhanced at elevated temperatures. These studies illustrate a complex intertwined relationship between ligation chemistry and assembly.

We have also exploited coralyne's ability to assemble DNA triplex structures in a ligation assay. Finally, an efficient solution to the protein-free coupling of monomer bases into oligomeric structures has been sought for decades. Our work with proflavine has shown that assembly and coupling of monomers into larger order structures can be successfully carried out, with minimal salt, at even submicromolar concentrations of nucleic acids.

6.2. EXPERIMENTAL PROCEDURES

6.2.1. Synthesis of Oligonucleotides

Substrate oligodeoxynucleotides were synthesized on an automated synthesizer using the phosphoramidite coupling chemistry. Substrates, 3'-phosphorothioate-(dT), 3'-phosphorothioate-d(TCT), and 3'-phosphorothioate- d(GCAGCGTCG) were synthesized according to the protocols described previously where the oxidation step normally used in the cycle is replaced by a sulfurization reaction. Iodo substrate, 5'-iodo-d(TTGT) was synthesized similarly to the 5'-iodo-$(dT)_4$ substrate. Modified substrate strands were HPLC purified and analyzed by mass spectrometry. Monomers with 2-methylimidazole phosphate activation chemistry were prepared according to published protocols and characterized using TLC [36]. All DNA template strand oligonucleotides were purchased from IDT and desalted by passage over a standard desalting column. Stock solutions of oligonucleotides were prepared by resuspending freeze-dried purified samples in deionized H_2O. Oligonucleotide concentrations were determined spectrophotometrically.

6.2.2. Radiolabeling of Substrate Strands

All 3'-phosphorothioate substrates were radioactively labelled with ^{32}P-phosphate at the 5'-end according to protocols described previously in this thesis.

6.2.3. Ligand Stock Solutions

Proflavine hemisulfate and coralyne chloride stock solutions were purchased from Sigma.

6.2.4. Template-Directed Ligation Assay

Reactions were carried out in a solution containing 10 mM Tris buffer (pH 8.2), and 100 mM 2-mercaptoethanol. Ligation reactions were stopped by plunging reaction test tubes into liquid nitrogen and freeze-drying. Concentrations of nucleic acids and small molecules for the specific reactions are listed in figure captions. All reactions using 2-MeImp activated monomers were incubated at 4ºC for 24 h either in the present of high salts (1.2 M Na^{+} and 0.2 M Mg^{++}) or low salt (10 mM Tris, 10 mM Na^{+}).

6.2.5. Product Analysis and Quantification

Freeze-dried reaction samples were resuspended in 10 μl of a 7 M urea solution and loaded on to a denaturing, 30% polyacrylamide gel (19:1::acrylamide:bisacrylamide). Gel electrophoresis and product quantification was carried out as described previously.

6.3. RESULTS AND DISCUSSION

6.3.1. Condensation of Monomers Using 2-Methylimidazole Phosphate Activation Chemistry

We have synthesized activated monomers using the following monophosphates: AMP, CMP, GMP, UMP. In Figure 6.1, 2-methylimidazole phosphate adenosine is shown.

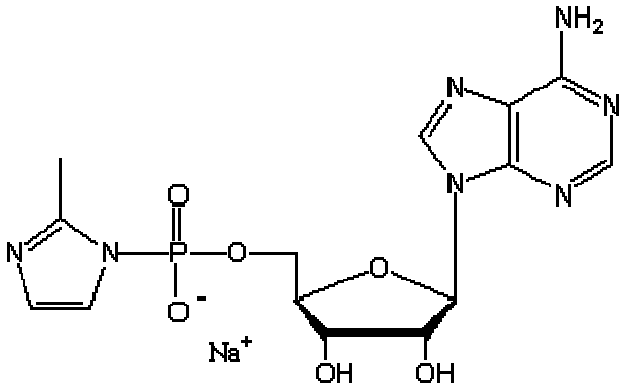


Figure 6.1. Chemical structure of 2-MeImp Phosphate Adenosine.

Attack of a sugar 2'-OH or 3'-OH group on the 5'- carbon group bearing the imidazole leaving moiety leads to condensation reactions. These reactions rely on the presence of molar quantities of NaCl and $MgCl_2$ in the reaction mixtures. We have carried out reactions where we attempt to condense 2-MeImp Adenosine on radiolabeled $(rA)_7$ or $(rA)_{16}$ primers (Figure 6.2). Poly (rU) and $(rU)_{16}$ templates were used. The sugar group on primer strands function as the nucleophile. Condensation reactions using Orgel's conditions have been successfully repeated. Lanes 2 and 7 in figure 6.2 show stepwise monomer addition of activated AMP on a $(rA)_7$ and $(rA)_{16}$ primer strand using Poly (U) templates.

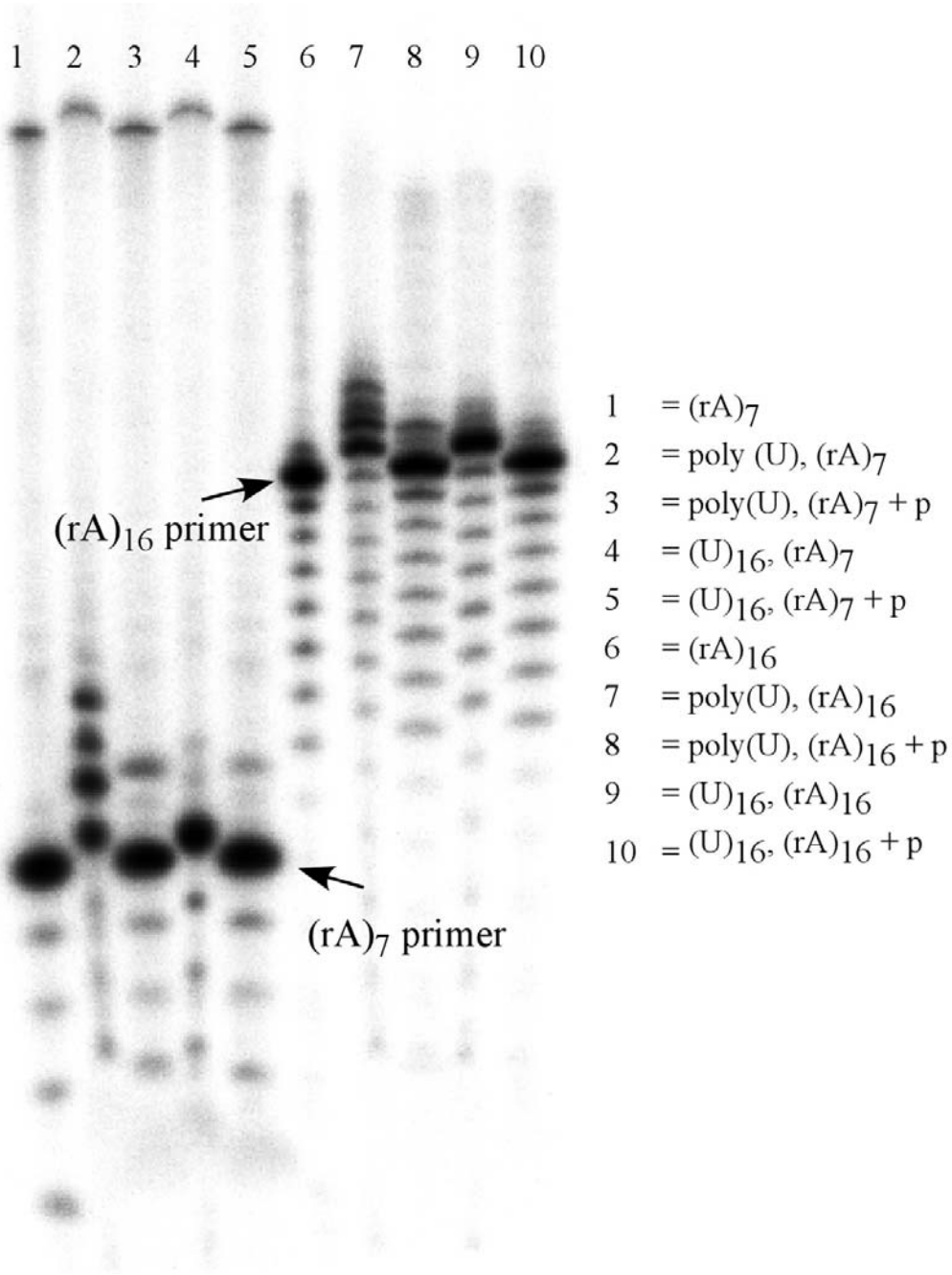


Figure 6.2. Denaturing PAGE analysis illustrating the condensation of 2-MeImp adenosine on radiolabeled $(rA)_n$ primers alongside $(rU)_n$ templates. Concentration of primers was 2 μM per strand, template = 1 μM per strand, activated AMP was 10 mM. All samples contained high salt (see methods) except marker lanes (1, 6) and lanes marked (+ p) which contained 100 μM proflavine.

Reactions are less favorable when using shorter $(rU)_{16}$ templates (lanes 4 and 9). Longer templates can potentially be more efficient due to greater base preorganization provided by intrastrand stacking [37]. It is not surprising to observe minimal chain growth for the reaction in lane 9 because the template strand is equal in length to the primer strand and thus would not allow chain extension. Reactions with proflavine were carried out in the presence of 10 mM Tris and 10 mM Na^+ (lanes 3, 5, 8, 10). It appears that reactions in the presence of proflavine are limited to single addition and are lower in efficiency than the reactions in the presence of high salts. The difference in mobility between the two types of samples (high salt and low salt) is also due to the different ionic strengths of the loaded samples.

Activated AMP reactions have been carried out in the presence of ethidium bromide, coralyne and BEPI. Ethidium bromide is a classical duplex intercalator [69] whereas coralyne and BePI are known to stabilize DNA triplexes [72, 86, 88, 103, 115, 163-165]. In addition to the effect of small molecules, Figure 6.3 illustrates the stringent requirements of cation type and concentration on condensation reactions in the absence of small molecules.

Lanes 2 – 6 in Figure 6.3 show that the presence of Mg^{++} is necessary for efficient chain growth polymerization of activated monomers. In the presence of 1.2 M NaCl, the reactions exhibit poor yields (lane 6). Reactions with small molecules show a concentration dependent decline in coupling and negligible enhancement in comparison to a reaction without small molecules (Figure 6.3, lane 8). A close inspection of reactions carried out with coralyne show that the loss in radiolabeled primer band occurs as a function of increasing concentration of coralyne. This result could be due to the

formation of coralyne mediated self structure of $(rA)_7$ and the retardation of the $(rA)_7$ primer band in the wells [86-88]. However, it is an intriguing phenomenon because gels run with 7M urea are expected to denature any self-structure formation.

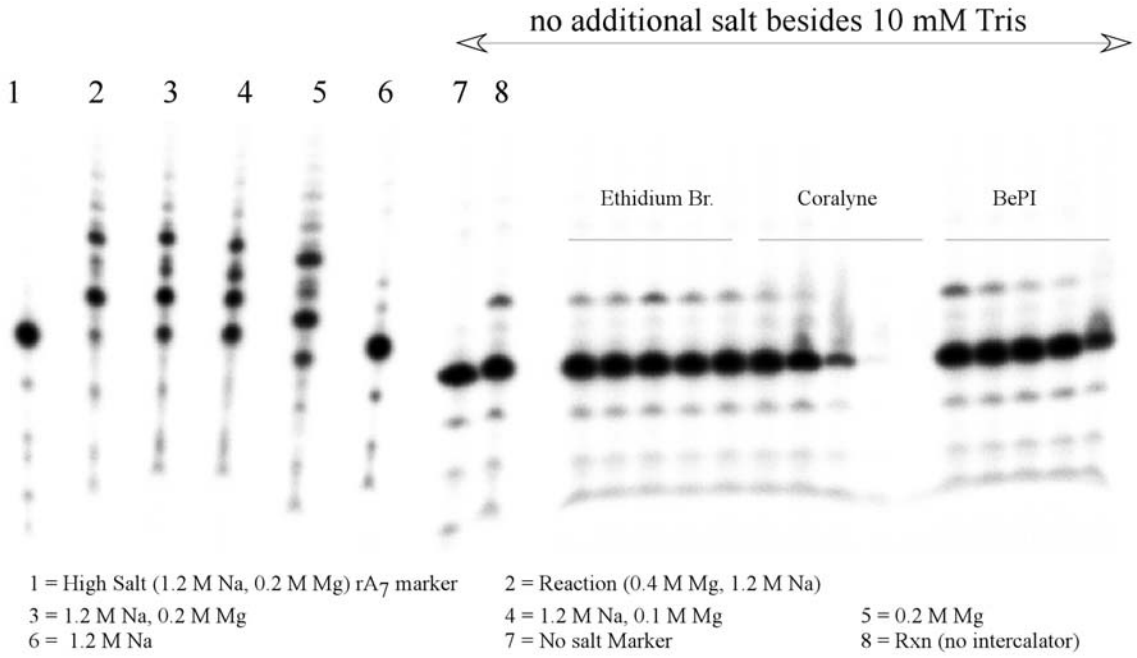


Figure 6.3. Effect of small molecules on 2-MeImp condensation reactions.
Lane designations are indicated below the figure. All samples contained 1 μM per strand of Poly (rU) template and $(rA)_7$ primer in addition to 10 mM activated AMP. Concentration of small molecules in μM (left to right): 50, 100, 250, 500, 1000.

We have looked at the effect of having a DNA:RNA hybrid structure on the monomer coupling reactions in the presence of proflavine and daunomycin. Work done by Chaires and coworkers has shown that daunomycin binds to a hybrid Poly(rA)•Poly(dT) with a greater affinity than Poly(rU)•Poly(rA) duplex [72]. For these experiments, radiolabeled $(rA)_7$ was used as the primer strand whereas $(dT)_{16}$ and $(dT)_{32}$ strands were used as the templates (Figure 6.4).

The lower mobility band in Figure 6.4 is the radiolabeled primer $(rA)_7$. We did not observe significant changes in coupling efficiency (quantity of product formation or the number of added monomers) in our trials with proflavine and daunomycin. Gel lanes in Figure 6.4 illustrate that we actually observe a decline in product formation as a function of increasing small molecule concentration.

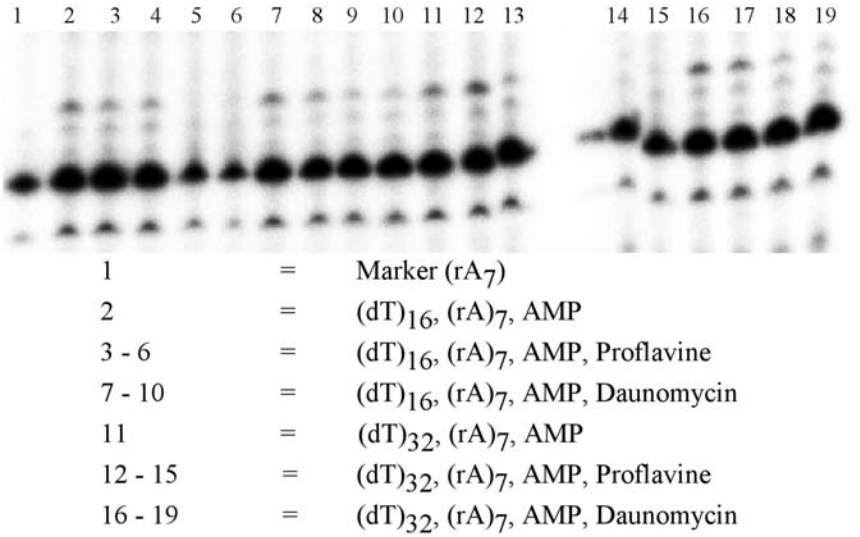


Figure 6.4. Gel electrophoresis illustrating 2-MeImp AMP condensation reactions in the presence of DNA templates. Concentrations of primer and template strands were 1 μM in strand whereas 10 mM 2-MeImp AMP was added to the reaction mixtures. Concentrations of small molecules in μM (left to right): 50, 100, 500, 1000. Reactions were carried out at 4ºC the presence of 10 mM Tris, pH 8.0.

Our initial results indicate that small molecule mediated assembly of nucleic acids is not compatible with the 2-methylimidazole phosphate activation chemistry. Our results show minimal enhancement at low concentrations of small molecules and a significant decline in product formation at higher concentrations of small molecules. In addition, we observe a single coupling event instead of chain growth polymerization in the presence of low salts which tends to suggest that condensation of activated monomers into higher order structures is dependent upon the presence of high quantities of NaCl and $MgCl_2$. It

may be possible that other chemical activations, besides phosphorothioate-iodo chemistry, are compatible with small molecule mediated assembly of nucleic acids.

6.3.2. Proflavine-Mediated Assembly of a Ligation Complex with Mixed Base Pairs

Proflavine mediated assembly and ligation of nucleic acids has been limited to homo (dA) templates and homo (dT) substrates. We wanted to test if proflavine mediated assembly and ligation of nucleic acids can be extended to mixed base pairing systems. We again designed an unstable system that does not assemble into a stable ligation complex in the absence of small molecules (Figure 6.5). Substrates I, 3'-phosphorothioate-d(TCT), and substrate II, 5'-iodo-d(TTGT), are aligned alongside d(ACAAAGA) template (lane 3, Figure 6.6).

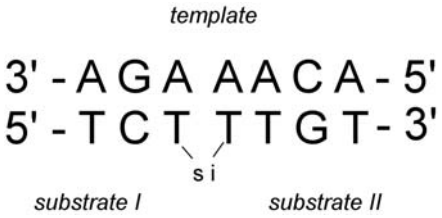


Figure 6.5. Schematic representation of a ligation active complex with mixed bases. 3'-phosphorothioate-d(TCT) and 5'-iodo-d(TTGT) are aligned alongside 5'-d(ACAAAGA) template strand.

Template-directed ligation experiments show that coupling of substrates is dependent upon the presence of proflavine (Figure 6.6). A competition reaction where two substrates, 3'-phosphorothioate-d(TTT) and 3'-phosphorothioate-d(TCT) were added together along with 5'-iodo-d(TTGT) in the presence of a 5'-ACAAAAA, shows that the 5'-iodo-d(TTGT) selectively couples with 3'-phosphorothioate-d(TTT). This is indicated

by a shift in the product band (lane 4 vs. lane 3, Figure 6.6). The data presented here demonstrates that small molecule mediated assembly of nucleic acids is not limited to strands containing only A•T base pairs.

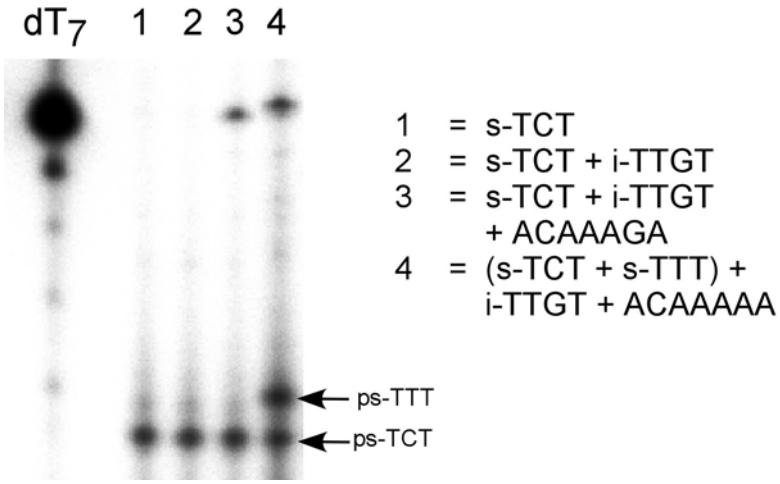


Figure 6.6. Denaturing PAGE illustrating assembly of a ligation complex with mixed base pairs. Reaction mixtures contained 10 mM Tris (pH 8.0), substrates and template 1 μM per strand each and proflavine at a concentration of 125 μM.

6.3.3. Experiments with a Stable Ligation System

6.3.3.1. Proflavine Adversely Affects a Stable Ligation Complex

It has been established that binding of proflavine leads to the assembly of a ligation active complex. However, proflavine may also affect the phosphorothioate coupling chemistry. Assembly is necessary for coupling of substrates to take place but the process and the solution conditions that lead to the assembly may also affect the chemical step of ligation. To test this possibility, we have designed a stable Watson-Crick paired system that does not require proflavine for assembly. A schematic representation of this test system is shown in Figure 6.7. Substrates spontaneously base pair with the complementary template strand and ligation takes place in a matter of seconds.

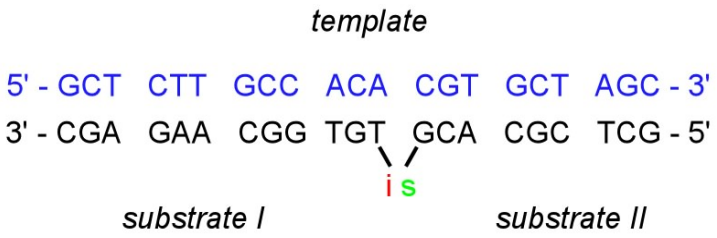


Figure 6.7. Schematic representation of a stable ligation active system. 5'-iodo modified substrate I and 3'-phosphorothioate modified substrate II are aligned alongside the complementary template.

Addition of proflavine to a stable ligation system leads to a decline in ligation activity. In Figure 6.8, it is shown that addition of proflavine up to a concentration of 60 µM does not affect the coupling of substrates substantially. However, further addition of proflavine is deleterious to coupling chemistry. It may be possible that at higher concentration of proflavine, proflavine interacts with the chemical modifications on the substrate strands. We cannot rule out the possibility that assembly is affected in an adverse manner as a function of increasing proflavine concentration.

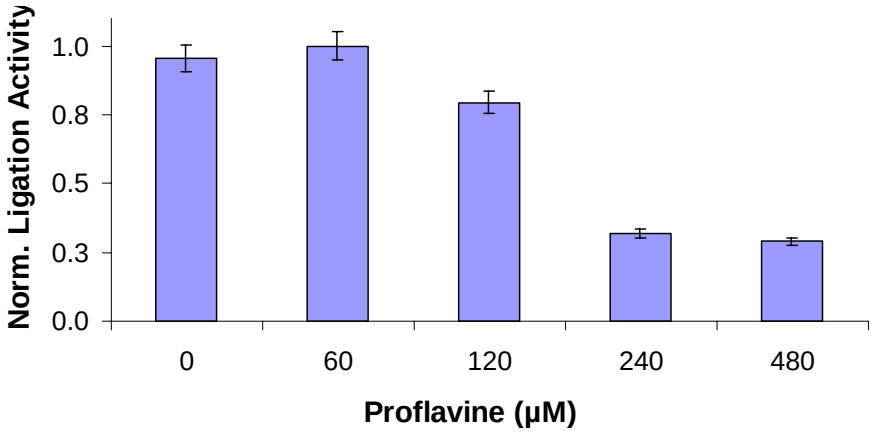


Figure 6.8. Quantitative analysis on the effect of proflavine on the coupling chemistry of a stable ligation system. Samples contained 10 mM Tris, 120 µM in base pair of ligation complex. Samples were incubated at 22ºC for 1h.

6.3.3.2. Coupling Chemistry is Enhanced at Elevated Temperatures

Reaction rates generally increase as a function of increasing temperature, as there is greater kinetic energy available to jump a transition state barrier. However, we also know that ligation active complex is more stable at reduced temperatures. Therefore, a strong interplay must exist between chemistry and assembly. The ligation system discussed above is stable up to 47ºC which allows us to study the effects of temperature below this range without affecting assembly.

We have measured reaction rates for a stable system at 4, 22, and 30ºC. A plot of reaction rates (ln k) as a function of temperature in Kelvin (1/T) shows a linear relationship. The slope of the graph allows direct determination of the activation energy for this reaction according to the Arrhenius equation.

$$k = A^{*}e^{-Ea/RT} \tag{6.1}$$

where k is the rate coefficient, A is a constant, E_a is the activation energy, R is the universal gas constant, and T is the temperature (in degrees Kelvin).

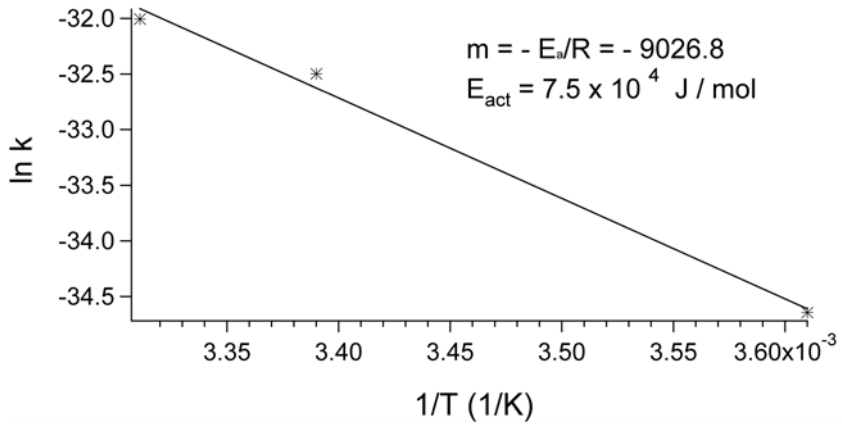


Figure 6.9. Plot of ln K vs. 1/T for a stable ligation system. Slope of the graph yields activation energy for this reaction. Reaction rates were determined from gel analysis.

The above results demonstrate that the rate of phosporothioate chemical step is elevated at higher temperatures. This data has increased our appreciation for the importance of coupling chemistry (bond formation rates) in the template-directed ligation experiments.

6.3.4. Coralyne Mediated Assembly of a Triple Stranded Ligation Active Complex

We have previously shown that coralyne mediates the assembly of stable DNA base triplets. We have also demonstrated that proflavine interaction with nucleic acids can increase the template-directed ligation rate of short oligonucleotides in a duplex system by three-orders of magnitude. Taking clues from the above two experimental demonstrations showing assembly as well as ligation, we designed a hairpin nucleic acid system which is dependent upon triplex formation for assembly and ligation (Figure 6.10). It has been shown that a GAA loop lends enormous stability (76ºC) to nucleic acid structures with only two GC pairs in the stem region [166].

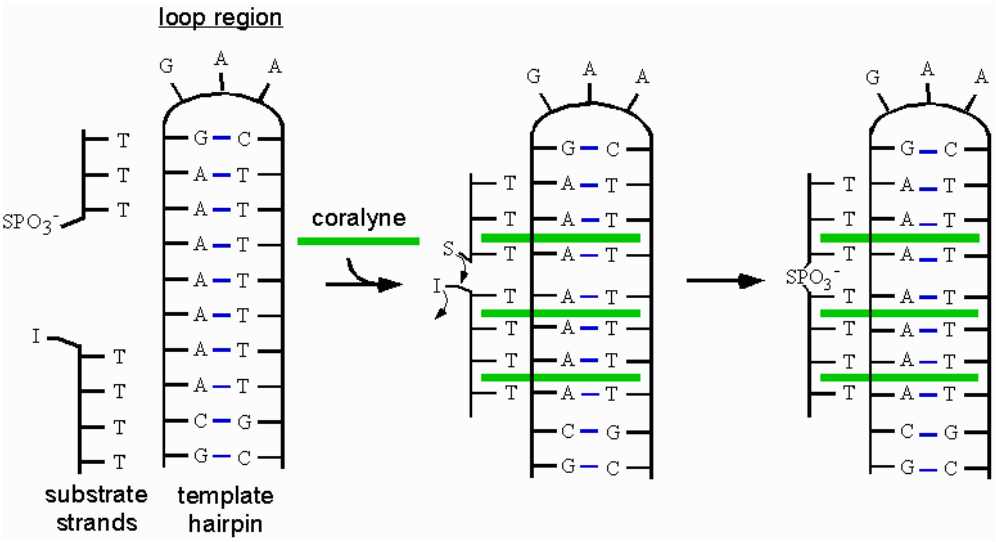
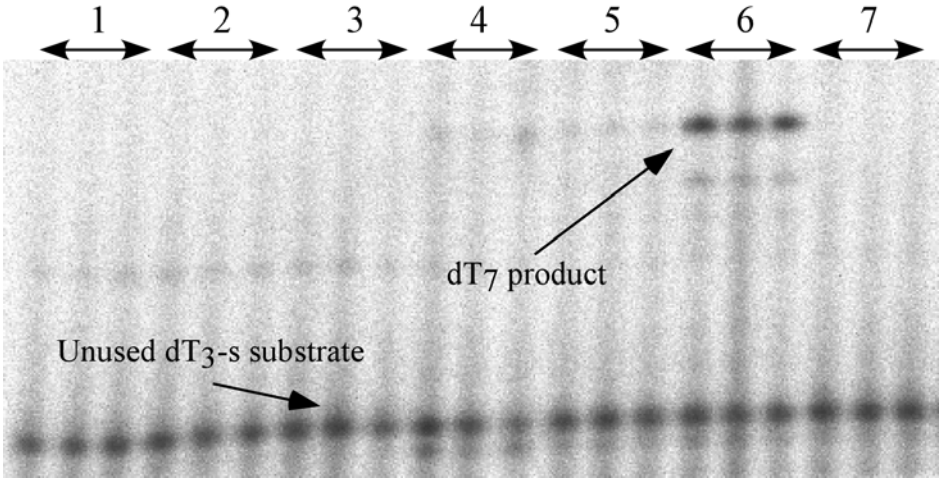


Figure 6.10. Schematic representation of a coralyne dependent triplex ligation active complex. Addition of coralyne recruits the substrate strands alongside the hairpin template into a triplex structure where coupling of substrates takes place.



1 = dT_3 *2 = $dT_3 + dT_4$*
3 = $dT_3 + dT_4$ + HP *4 = $dT_3 + dT_4$ + Coralyne*
5 = $dT_3 + dT_4$ + Proflavine *6 = $dT_3 + dT_4$ + HP + Coralyne*
7 = $dT_3 + dT_4$ + HP + Proflavine

Figure 6.11. Denaturing PAGE illustrating assembly of a triplex ligation active complex. Samples contained 10 mM Tris (pH 8.0), 1 μM per strand nucleic acids and 125 μM coralyne or proflavine. Solutions were incubated at 4ºC for 24 h. Samples were run in triplicate.

Data presented in Figure 6.11 shows that assembly and ligation is dependent upon the presence of coralyne (lane 6, Figure 6.11). Addition of proflavine to a reaction mixture containing substrates and a hairpin template yields no ligation product (lane 7, Figure 6.11). Our work has shown that simple switching of the template strand to a hairpin which affords the formation of DNA triplexes leads to the assembly of a coralyne-dependent ligation system. These results show the remarkable tunability of the small molecule based ligation of nucleic acids.

6.3.5. Protein Free Assembly and Coupling of Monomers

Non-enzymatic condensation of monomers represents the Holy Grail in the origin of nucleic acids. The RNA World hypothesis is dependent upon the presence of RNA or

RNA-like polymers [12, 13, 167]. An efficient solution to the coupling of monomers is still lacking. We have designed a template-based system where one of the substrate molecules is a 3-'phosphorothioate-(dT). A schematic representation of this test system is shown in Figure 6.12.

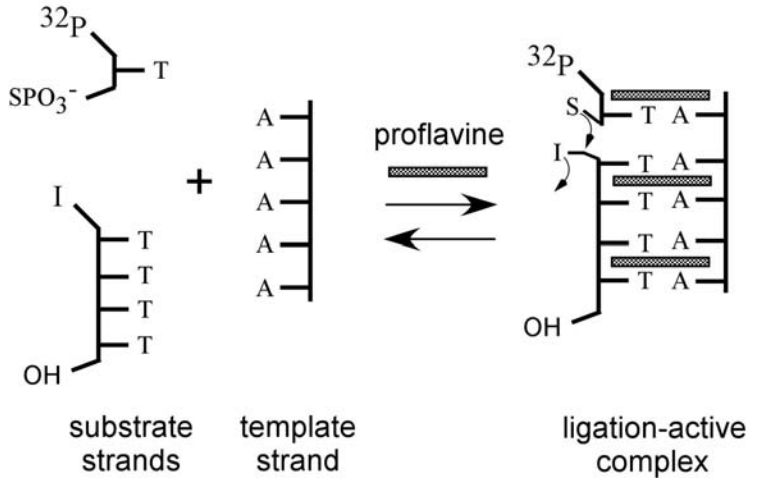


Figure 6.12. Schematic representation of a monomer based ligation test system. Substrates 3'-ps-(dT), 5'-iodo-$(dT)_4$ and template strand $(dA)_5$ were present in solution. Addition of proflavine assembles substrates alongside template strands into a ligation active complex.

An increase in the intensity of a band corresponding to the mobility of a $(dT)_5$ marker band is observed as a function of increasing proflavine concentration (Figure 6.13). Upon closer observation, a minor product is also observed which is a $(dT)_4$ band. Modified substrate, 5'-iodo-$(dT)_4$ contains a truncation product which is the 5'-iodo-$(dT)_3$. This truncation product couples to the monomer and yields a product with a gel mobility of a $(dT)_4$ oligonucleotide.

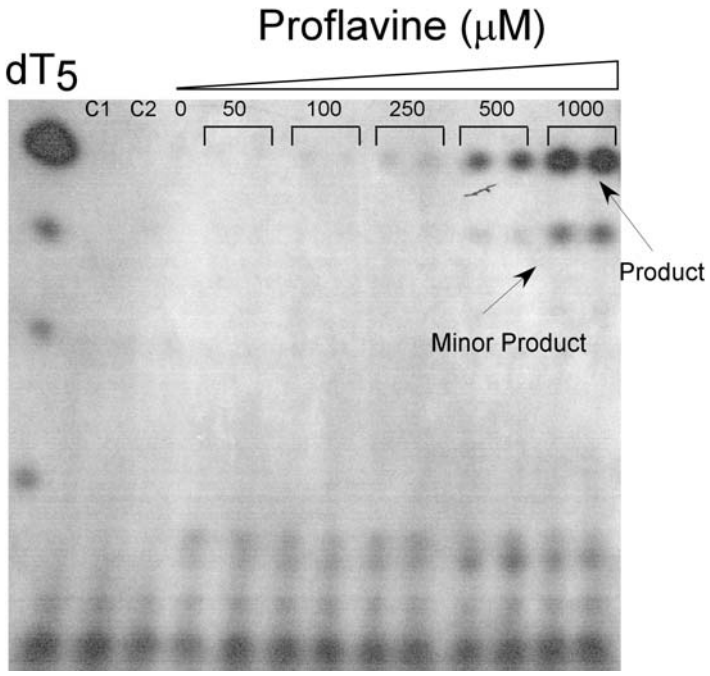


Figure 6.13. Denaturing PAGE analysis illustrating monomer coupling.
Modified monomer couples with a tetramer substrate and yields a $(dT)_5$ product band which increases in intensity as a function of increasing proflavine concentration. Solution conditions: 10mM Tris (pH 8.0), nucleic acids were 960 nM in strand each. Lane C1: 3'-phosphorothioate-dT substrate, C2: 3'-phosphorothioate-dT + 5'-iodo-$(dT)_4$. All other lanes contained substrates 3'-phosphorothioate-dT & 5'-iodo-$(dT)_4$ and template $(dA)_5$. Proflavine concentration in μM (left to right): 0, 50, 100, 250, 500, 1000. Lanes containing proflavine were loaded in duplicate.

In the model in Figure 6.11, we have shown that a proflavine molecule caps the ligation complex terminus by stacking on top of the modified monomer base. This capping effect may be playing a very crucial role in the assembly and ligation in this system shown here. It has been shown by Richert *et al*. that a molecule, whose shape mimics base pairs, when tethered to nucleic acid termini, functions as a potent stabilizing agent by capping nucleic acid termini and reducing the end fraying effect [168].

Overall, our results show that small molecule binding to nucleic acids is a very powerful, robust, and versatile technique for assembly of nucleic acid secondary structures.

6.3.6. HPLC Purification and Mass Spectrometry Analysis of the Modified Substrates

6.3.6.1. 3'-phosphorothioate-d(GCAGCGTCG)

Figure 6.14 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time for the purification of 3'-phosphorothioate-d(GCAGCGTCG) oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 6% acetonitrile, 50 min = 25% acetonitrile. Flow rate was 4.6 ml/min.

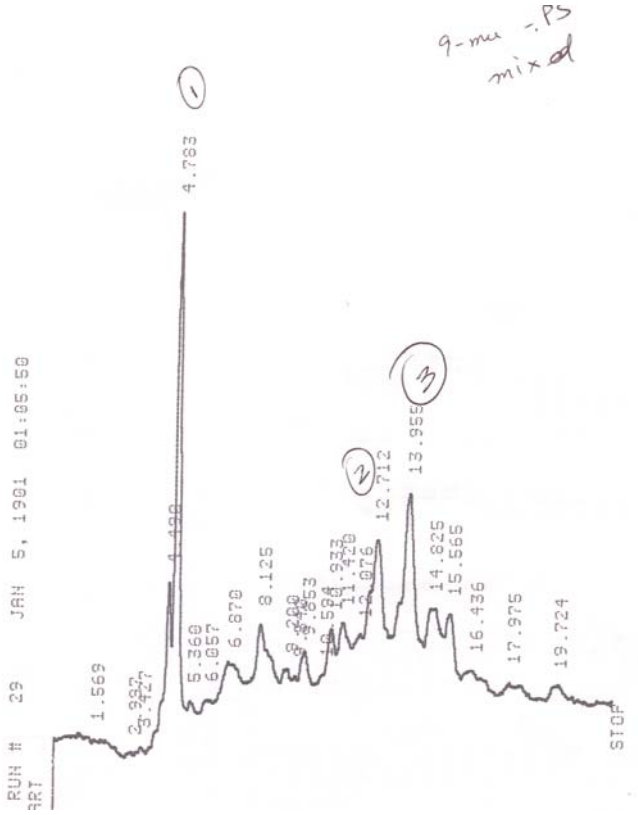


Figure 6.14. HPLC purification of 3'-phosphorothioate-d(GCAGCGTCG).
Chromatograph of a crude sample is shown. Three fractions were collected where fraction # 3 was designated as major peak. All three fractions were analyzed by mass spectrometry.

Fraction # 3 from HPLC analysis was confirmed by mass spectrometry as the product of interest. Mass spectrum of a purified 3'-phosphorothioate-d(GCAGCGTCG) sample is shown in Figure 6.15. Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 2834.8 g/mol. M-H peak is observed at 2834.0 g/mol.

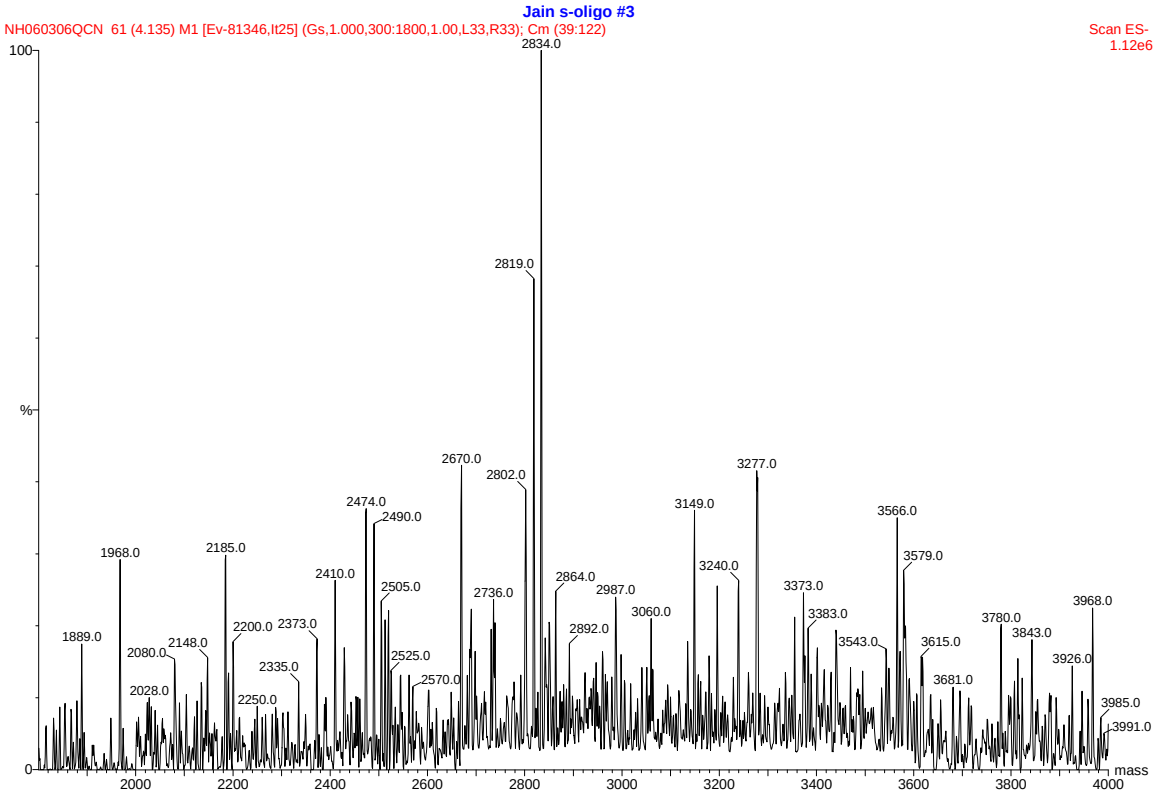


Figure 6.15. Mass spectrum of a 3'-phosphorothioate-d(GCAGCGTCG). Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 2834.8 g/mol. M-H peak is observed at 2834.0 g/mol.

6.3.6.2. 5'-iodo-(TGTGGCAAGAGC)

Figure 6.16 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time for the purification of 5'-iodo-(TGTGGCAAGAGC) oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 8% acetonitrile, 50 min = 20% acetonitrile. Flow rate was 4.6 ml/min.

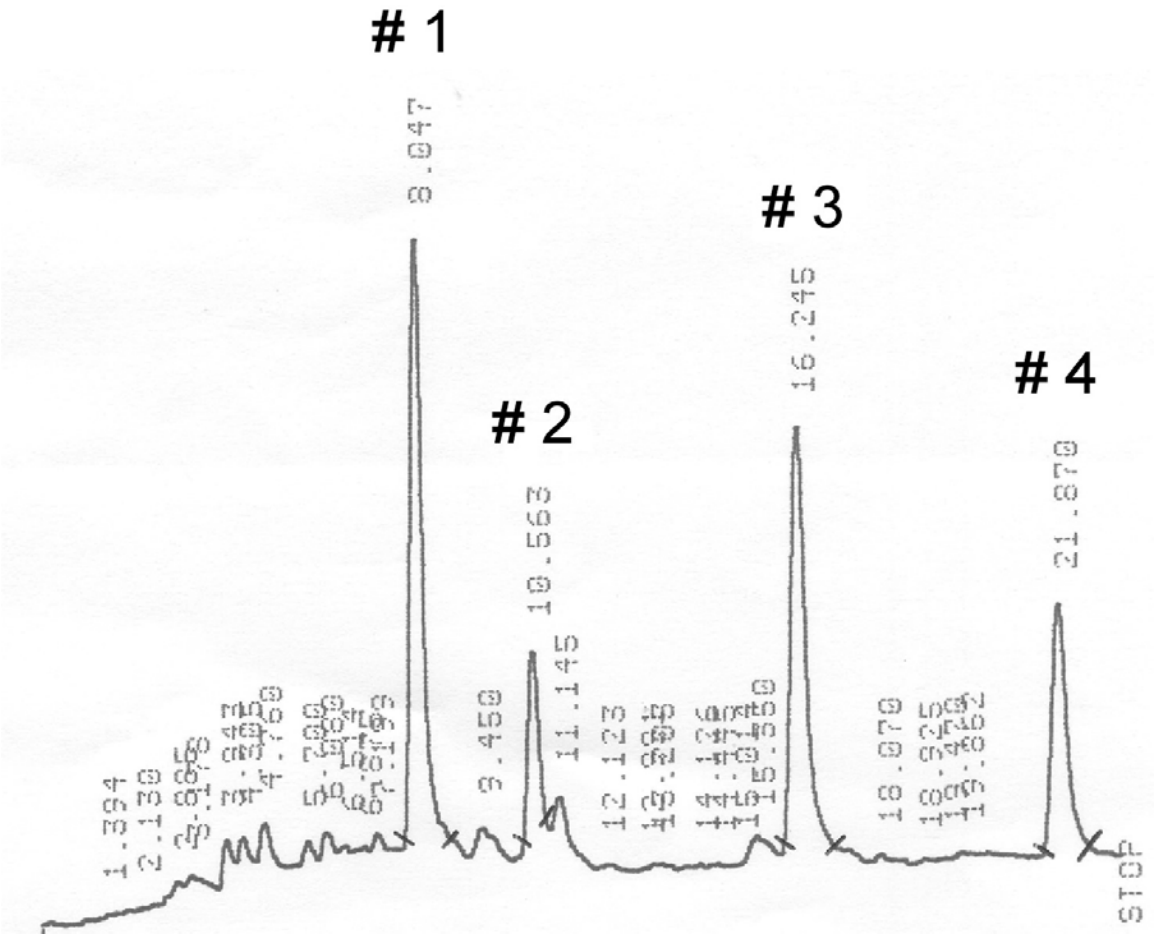


Figure 6.16. HPLC purification of a 5'-iodo-(TGTGGCAAGAGC) oligonucleotide. Chromatograph of a crude sample is shown. Four major peaks were collected and submitted for mass spectrometry analysis.

Mass spectrum of a purified 5'-iodo-(TGTGGCAAGAGC) oligonucleotide sample is shown in Figure 6.17. Peak 4 from the HPLC purification was identified as the compound of interest. Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 3820.4 g/mol. M – H peak is observed at 3819.0 g/mol which is the compound of interest. We also observe a significant peak at 3751 g/mol and 3692 g/mol. These observations are surprising as we isolate well separated peaks in the HPLC chromatogram. Loss of the 5'-group in its entirety would yield the peak at 3692 g/mol. We are not able to confirm the identity of the peak at 3751 g/mol. Previous experiments and MS analysis has shown that the 5'-iodo group undergoes degradation over time.

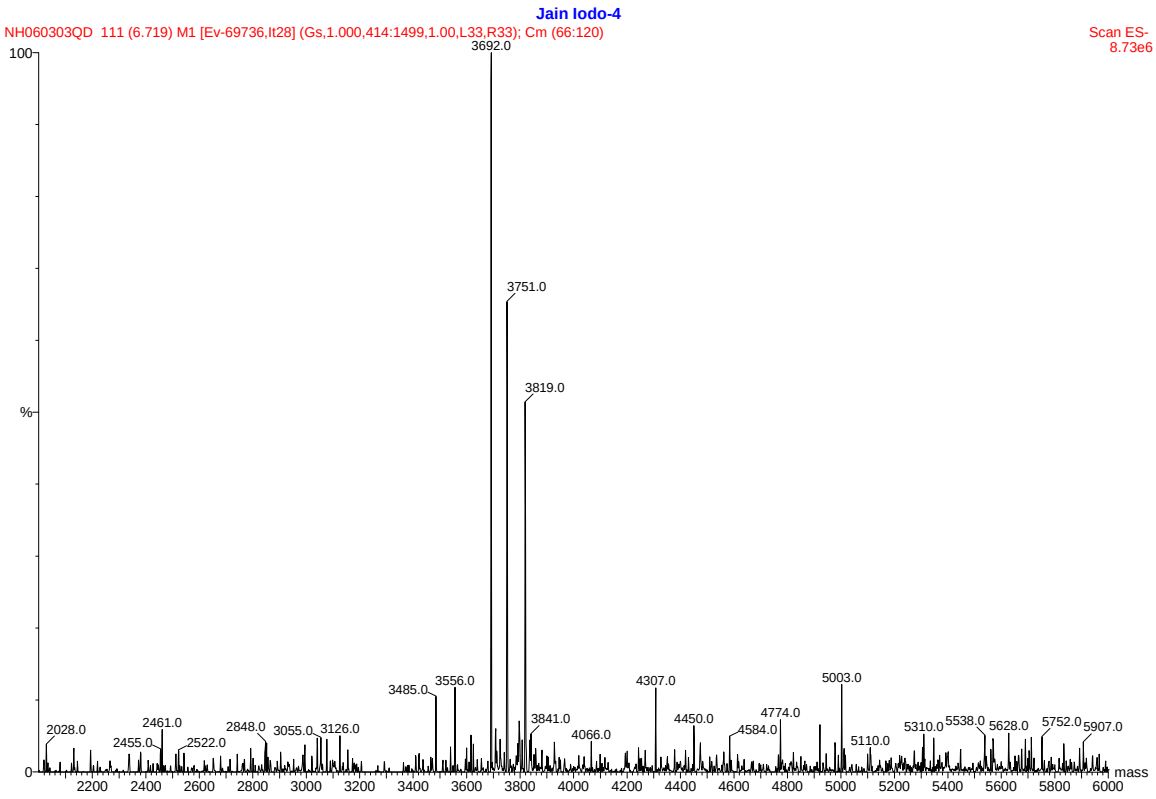


Figure 6.17. Mass Spectrum of a 5'-iodo-(TGTGGCAAGAGC) oligonucleotide.

6.3.6.3. 3'-phosphorothioate-d(TCT)

Figure 6.18 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time for the purification of 3'-phosphorothioate-d(TCT) oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 6% acetonitrile, 50 min = 20% acetonitrile. Flow rate was 4.6 ml/min.

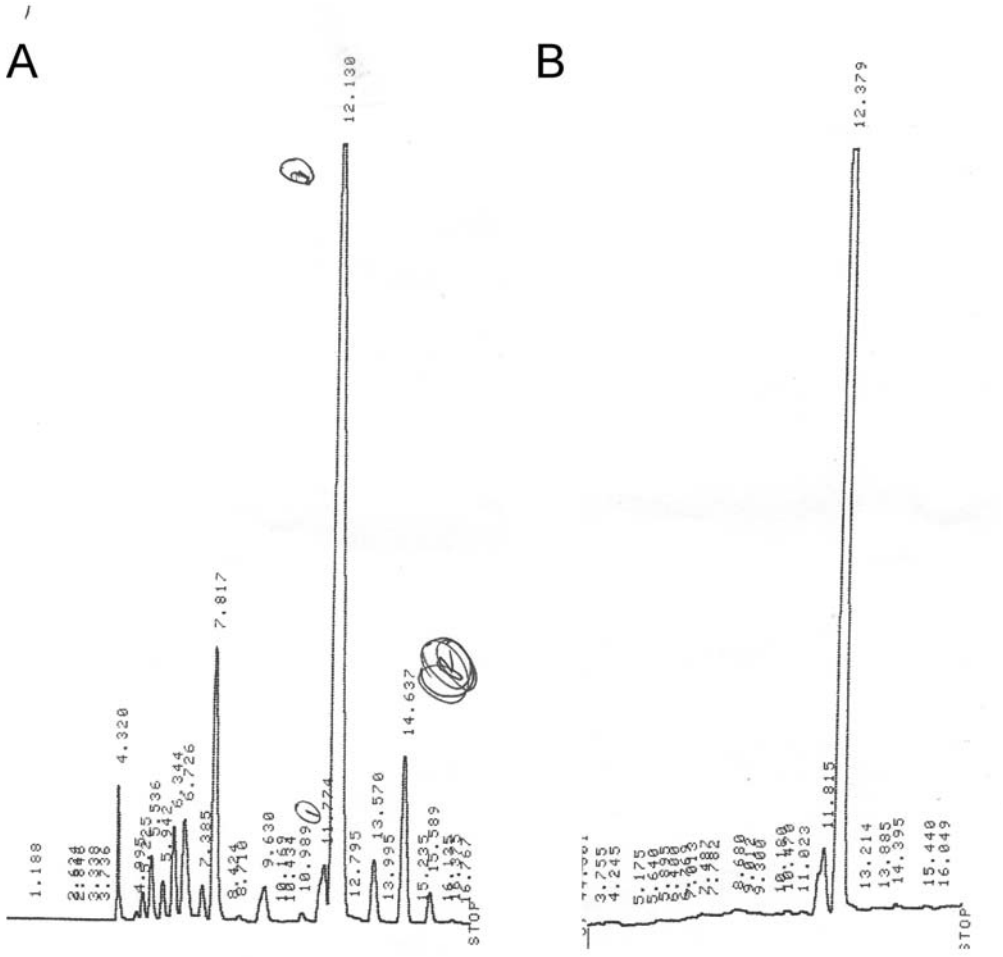


Figure 6.18. HPLC purification of 3'-phosphorothioate-d(TCT). (A) HPLC chromatograph of a crude sample is shown. Only one major peak was present which was collected. (B) Fraction # 1 (major) was run on the column a second time to confirm purity.

6.3.6.4. 3'-phosphorothioate, 5'-phosphate-d(TCT)

Figure 6.19 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time for the purification of 3'-phosphorothioate, 5'-phosphate-d(TCT) oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 6% acetonitrile, 50 min = 20% acetonitrile. Flow rate was 4.6 ml/min.

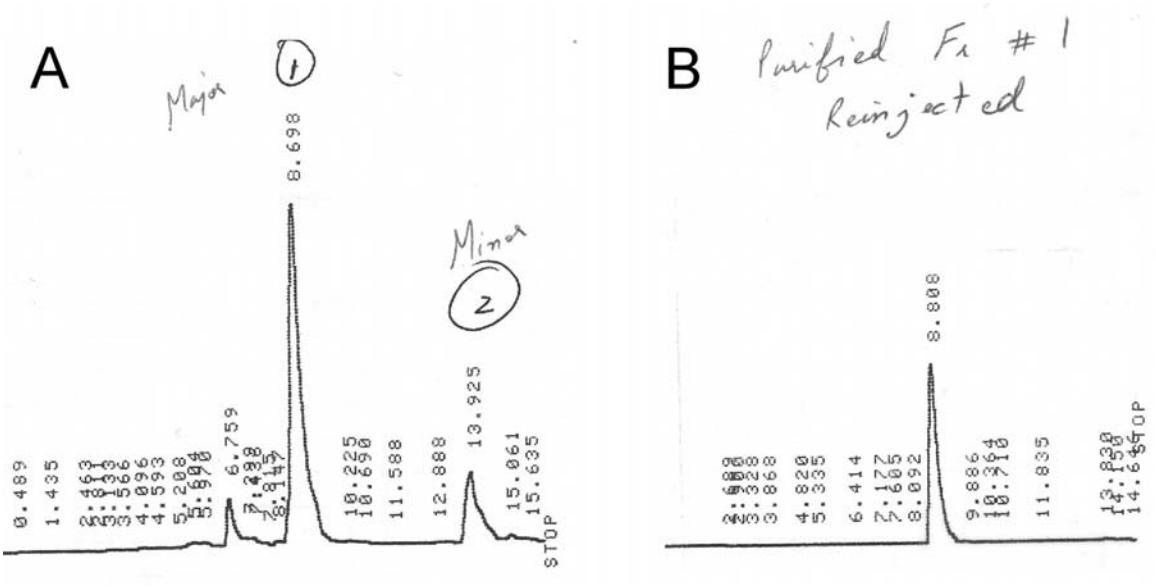


Figure 6.19. HPLC purification of 3'-phosphorothioate, 5'-phosphate-d(TCT) oligonucleotide. (A) Chromatograph of a crude sample is shown. Two fractions (#1 and #2) were collected. (B) Fraction # 1 (major) was run on the column a second time to confirm purity. Sample identities were confirmed by mass spectrometry analysis.

Mass spectrum of a purified 3'-phosphorothioate, 5'-phosphate-d(TCT) oligonucleotide sample is shown in Figure 6.20. Spectrum was collected in ESI negative ion mode. Theoretical mass of the oligonucleotide is 1010.8 g/mol. Peak A observed at 1009.8 g/mol is the M – H peak which is the compound of interest. We also observe M - 2H (504.6 g/mol) which indicates doubly charged species in the sample.

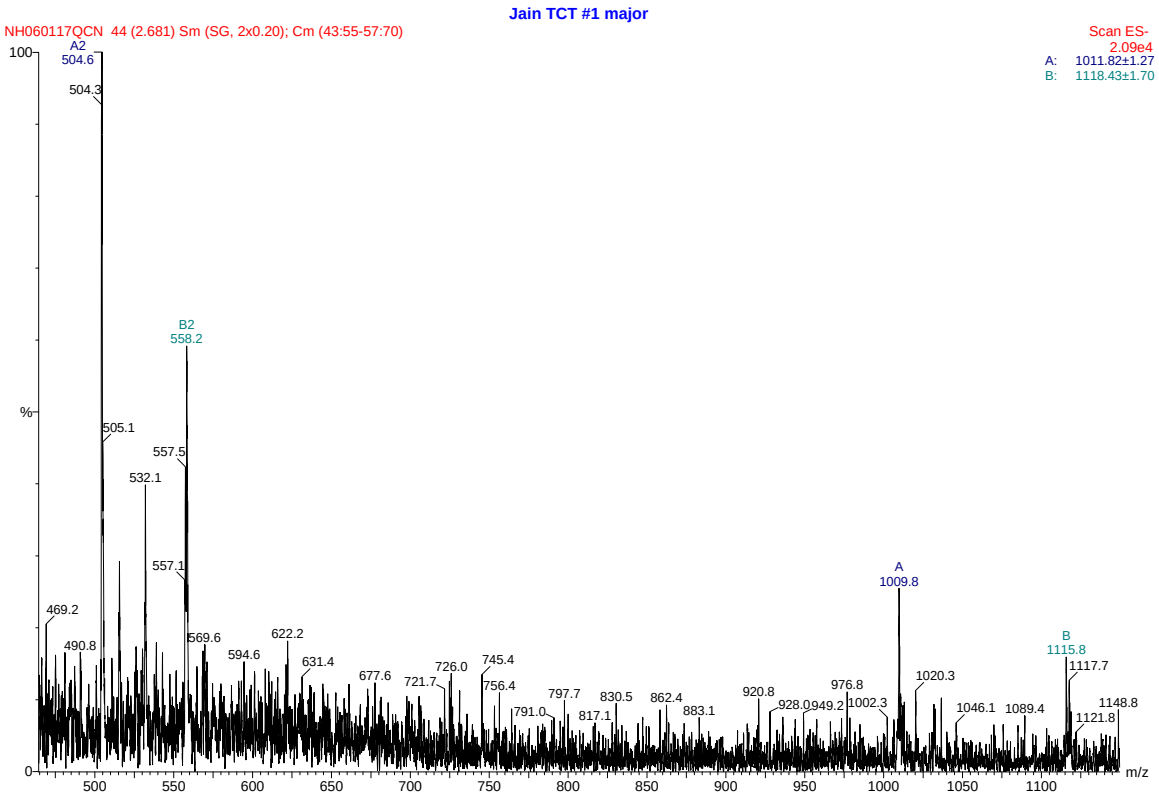


Figure 6.20. Mass Spectrum of a 3'-phosphorothioate, 5'-phosphate-d(TCT) oligonucleotide.

6.3.6.5. 5'-iodo-d(TTGT)

Figure 6.21 shows the HPLC chromatograph (Abs monitored at λ 269 nm) as a function of time for the purification of 5'-iodo-d(TTGT) oligonucleotide. Purification was achieved on a Phenomenex ODS semi preparative column (10 x 250 mm, 5 μ). Aqueous mobile phase A was 0.1 M TEAA and organic elution solvent (B) was acetonitrile. Gradient conditions were as follows: 0 min = 8% acetonitrile, 50 min = 25% acetonitrile. Flow rate was 4.6 ml/min.

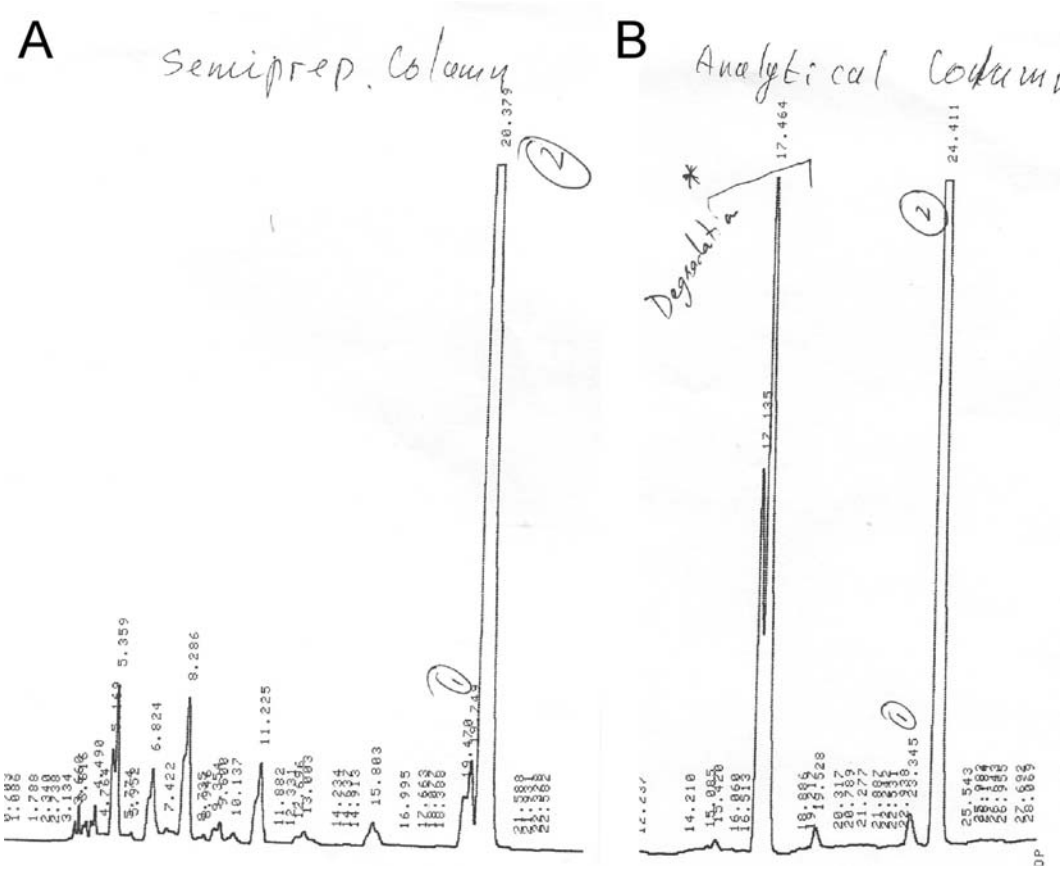


Figure 6.21. HPLC purification of 5'-iodo-d(TTGT). (A) Chromatograph of a crude sample is shown. Two fractions (#1 and #2) were collected. (B) Fraction # 2 (major) was run on the column a second time to confirm purity. Degradation of the purified sample is apparent as additional peaks at ~ 17 min retention time are observed.

6.4. CONCLUDING REMARKS AND FUTURE DIRECTIONS

Creation of an efficient self-replicating system remains an elusive goal. Small molecule binding to nucleic acids represents a powerful approach for assembly of nucleic acids. Demonstration of a system which achieves chain growth polymerization in the presence of small molecules would represent a major advancement in realizing self replication and autocatalysis. Work by us and others has shown that homo (dA) and homo (rA) strands can be self-structured in the presence of coralyne [72, 86-88, 116]. Data from our lab demonstrates that $(dA)_{16}$ oligonucleotides can be linked into long chain polymers with antiparallel strand orientation. Thus, an experimental system can be envisioned where the 3'- and 5'- ends of a $(dA)_n$ oligonucleotide are modified. Addition of coralyne can potentially form ladder-like structures and bring the reactive groups close in proximity to allow ligation and chain growth polymerization to take place. A similar type of chain growth system can also be designed to utilize proflavine's ability to assemble a Watson-Crick ligation active duplex. Detection in these systems can be afforded by seeding the reactions with a small quantity of substrate with modification at the 3'-end only where the 5'-end can be radiolabeled by a ^{32}P group. The discovery that ligation active complexes can be assembled with mismatched base pairs and more importantly, T•T base pairs, shows that additional experimental test systems can be designed which exploit non-natural pairings for assembly of nucleic acids. These studies can ultimately carve a path towards the design of artificial self-replication systems with pairing of like substrates and templates.

A thorough investigation of small molecules with different structures and binding modes also needs to be undertaken to explore their utility in assembly of unstable nucleic

acid structures. It is entirely possible that other small molecules, besides coralyne and proflavine, would promote ligation active assemblies in an experimental test system involving a different type of backbone linkage. For instance, minor groove binding agents facilitate in the non-enzymatic ligation of plasmid DNA [154]. In a similar fashion, bisintercalating molecules such as YOYO [169] may also prove useful in promotion of inter- or intrastrand ligations.

Our studies with coralyne and proflavine have revealed that small molecule binding to nucleic acids is significantly more complicated than we previously appreciated. The kinetics and thermodynamics of DNA secondary structure formation in the presence of small molecules can be a complex phenomenon. Part of this complexity arises from the fact that small molecules, such as proflavine and coralyne, have more than one binding mode to nucleic acids. Thorough understanding of additional binding modes of these molecules is crucial to understanding their true mode of action and mechanism. Nevertheless, our work has shown that small molecule binding to nucleic acids can be a very powerful and robust technique for stable assembly of DNA and RNA secondary structures. Continued research efforts in the area can open new frontiers in the area of protein-free condensation of monomers and design of nucleic acid self replicating systems.

REFERENCES

1. Avery, O., MacLeod, C. and McCarty, M. (1944). "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types." *J. Expt. Med.*, **79**: 137-157.

2. Hershey, A.D. and Chase, M. (1952). "Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage." *J. Gen. Phys.*, **36**(1): 39-56.

3. Chargaff, E. (1950). "Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation." *Experimentia*, **6**(6): 201-209.

4. Watson, J.D. and Crick, F.H.C. (1953). "Molecular Structure of Nucleic Acids - A Structure for Deoxyribose Nucleic Acid." *Nature*, **171**(4356): 737-738.

5. Bloomfield, V., Crothers, D.M. and Tinoco, I. *Nucleic Acids: Structure, Properties and Function*, ed. J. Stiefel. 2000, Sausalito: University Science Books. 794.

6. Miller, S.L. (1953). "A Production of Amino Acids under Possible Primitive Earth Conditions." *Science*, **117**(3046): 528-529.

7. Oro, J. (1961). "Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions." *Nature*, **191**(479): 1193.

8. Oro, J. and Kamat, S.S. (1961). "Amino-Acid Synthesis from Hydrogen Cyanide under Possible Primitive Earth Conditions." *Nature*, **190**(477): 442.

9. Robertson, M.P. and Miller, S.L. (1995). "An efficient prebiotic synthesis of cytosine and uracil [published erratum appears in Nature 1995 Sep 21;377(6546):257]." *Nature*, **375**(6534): 772-4.

10. Miyakawa, S. *et al.* (2002). "Prebiotic synthesis from CO atmospheres: Implications for the origins of life." *Proc. Nat. Acad. Sci. USA*, **99**(23): 14628-14631.

11. Orgel, L.E. (2004). "Prebiotic Adenine Revisited: Eutectics and Photochemistry." *Orig. Life Evol. Biosph.*, **34**: 361-369.

12. Gilbert, W. (1986). "The RNA World." *Nature*, **319**: 618.

13. Kruger, K. *et al.* (1982). "Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*." *Cell*, **31**: 147-157.

14. Joyce, G.F. (1987). "Nonenzymatic Template-Directed Synthesis of Informational Macromolecules." *Cold Spring Harbor Symp. Quant. Biol.*, **52**: 41-51.

15. Joyce, G.F. (1989). "RNA evolution and the origins of life." *Nature*, **338**: 217-224.

16. Doudna, J.A., Couture, S. and Szostak, J.W. (1991). "A Multisubunit Ribozyme that is a Catalyst of and Template for Complementary Strand RNA-Synthesis." *Science*, **251**(5001): 1605-1608.

17. Doudna, J.A. and Szostak, J.W. (1989). "RNA-Catalyzed Synthesis of Complementary-Strand RNA." *Nature*, **339**(6225): 519-522.

18. Uhlenbeck, O.C. (1987). "A Small Catalytic Oligoribonucleotide." *Nature*, **328**(6131): 596-600.

19. Bartel, D.P., *Re-Creating an RNA Replicase*, in *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World*, R.F. Gesteland and J.F. Atkins, Editors. 1999, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. p. 143-162.

20. Barrick, J.E. *et al.* (2004). "New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control." *Proc. Nat. Acad. Sci. USA.*, **101**(17): 6421-6426.

21. Mandal, M. and Breaker, R.R. (2004). "Adenine riboswitches and gene activation by disruption of a transcription terminator." *Nat. Struct. Mol. Biol.*, **11**(1): 29-35.

22. Mandal, M. *et al.* (2004). "A glycine-dependent riboswitch that uses cooperative binding to control gene expression." *Science*, **306**(5694): 275-279.

23. Nahvi, A. *et al.* (2002). "Genetic control by a metabolite binding mRNA." *Chem. Biol.*, **9**(9): 1043-1049.

24. Winkler, W., Nahvi, A. and Breaker, R.R. (2002). "Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression." *Nature*, **419**(6910): 952-956.

25. Winkler, W.C. *et al.* (2004). "Control of gene expression by a natural metabolite-responsive ribozyme." *Nature*, **428**(6980): 281-286.

26. Winkler, W.C. *et al.* (2003). "An mRNA structure that controls gene expression by binding S-adenosylmethionine." *Nat. Struct. Biol.*, **10**(9): 701-707.

27. Breaker, R.R. *et al.* (2003). "A common speed limit for RNA-cleaving ribozymes and deoxyribozymes." *RNA*, **9**(8): 949-957.

28. Carmi, N., Shultz, L.A. and Breaker, R.R. (1996). "In vitro selection of self-cleaving DNAs." *Chem. Biol.*, **3**(12): 1039-1046.

29. Li, Y.F. and Breaker, R.R. (1999). "Deoxyribozymes: new players in the ancient game of biocatalysis." *Curr. Opin. Struct. Biol.*, **9**(3): 315-323.

30. Li, Y.F. and Breaker, R.R. (1999). "Phosphorylating DNA with DNA." *Proc. Nat. Acad. Sci. USA*, **96**(6): 2746-2751.

31. Li, Y.F. and Breaker, R.R. (2001). "*In vitro* selection of kinase and ligase deoxyribozymes." *Methods*, **23**(2): 179-190.

32. Coppins, R.L. *et al.* (2005). "Synthesis of native 3 '-5 ' RNA linkages by deoxyribozymes." *Abstracts of Papers of the American Chemical Society*, **229**: U530-U530.

33. Flynn-Charlebois, A. *et al.* (2003). "Deoxyribozymes with 2 '-5 ' RNA ligase activity." *J. Am. Chem. Soc.*, **125**(9): 2444-2454.

34. Orgel, L.E. (1987). "Evolution of the genetic apparatus: a review." *Cold Spring Harbor Symposia on Quantitative Biology*, **LII**: 9-16.

35. Chen, C.B., Inoue, T. and Orgel, L.E. (1985). "Template-Directed Synthesis on Oligodeoxycytidylate and Polydeoxycytidylate Templates." *J. Mol. Biol.*, **181**(2): 271-279.

36. Inoue, T. and Orgel, L.E. (1982). "Oligomerization of (Guanosine 5'-phosphor)-2-methylimidazolide on Poly(C)." *J. Mol. Biol.*, **162**: 201-217.

37. Kozlov, I.A. and Orgel, L.E. (2000). "Nonenzymatic template-directed synthesis of RNA from monomers." *Mol. Biol.*, **34**(6): 781-789.

38. Weimann, R.L. *et al.* (1968). "Template-Directed Synthesis with Adenosine-5'-phosphorimidazolide." *Science*, **161**(839): 387.

39. Wu, T.F. and Orgel, L.E. (1992). "Nonenzymatic Template-Directed Synthesis On Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides." *J. Am. Chem. Soc.*, **114**(1): 317-322.

40. Kanaya, E. and Yanagawa, H. (1986). "Template-directed polymerization of oligoadenylates using cyanogen bromide." *Biochemistry*, **25**: 7423-7430.

41. Ferris, J.P. *et al.* (1996). "Synthesis of long prebiotic oligomers on mineral surfaces." *Nature*, **381**(6577): 59-61.

42. Ferris, J.P., Huang, C.H. and Hagan Jr., W.J. (1989). "N-Cyanoimidazole and dimidazole imine: water-soluble condensing agents for the formation of the phosphodiester bond." *Nucleos. Nucleot.*, **8**: 407-414.

43. Ertem, G. and Ferris, J.P. (1997). "Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. A possible bridge between the prebiotic and RNA worlds." *J. Am. Chem. Soc.*, **119**(31): 7197-7201.

44. Ertem, G. and Ferris, J.P. (1996). "Synthesis of RNA oligomers on heterogeneous templates." *Nature*, **379**(6562): 238-240.

45. Ferris, J.P. and Ertem, G. (1992). "Oligomerization of Ribonucleotides on Montmorillonite - Reaction of the 5'-Phosphorimidazolide of Adenosine." *Science*, **257**(5075): 1387-1389.

46. Gat, Y. and Lynn, D.G. (1998). "Reading DNA differently." *Biopolymers*, **48**(1): 19-28.

47. Goodwin, J.T. and Lynn, D.G. (1992). "Template-directed synthesis: Use of a reversible reaction." *J. Am. Chem. Soc.*, **114**: 9197-9198.

48. Leitzel, J.C. and Lynn, D.G. (2001). "Template-directed ligation: From DNA towards different versatile templates." *Chem. Rec.*, **1**(1): 53-62.

49. Li, Z.Y. *et al.* (2002). "DNA catalyzed polymerization." *J. Am. Chem. Soc.*, **124**(5): 746-747.

50. Ye, J.D., Gat, Y. and Lynn, D.G. (2000). "Catalyst for DNA ligation: Towards a two-stage replication cycle." *Angew. Chem. Int. Ed.*, **39**(20): 3641-3643.

51. Joyce, G.F. (1987). "Nonenzymatic template-directed synthesis of information macromolecules." *Cold Spring Harbor Symp. Quant. Biol.*, **52**: 41- 51.

52. Joyce, G.F. and Orgel, L.E., *Prospects for understanding the origin of the RNA world*, in *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World*, R.F. Gesteland and J.F. Atkins, Editors. 1999, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. p. 49-77.

53. von kiedrowski, G. (1986). "A Self-Replicating Hexadeoxynucleotide." *Angew. Chem. Int. Ed.*, **25**(10): 932-935.

54. Tjivikua, T., Ballester, P. and Rebek, J. (1990). "A Self-Replicating System." *J. Am. Chem. Soc.*, **112**: 1249-1250.

55. Li, T. and Nicolaou, K.C. (1994). "Chemical Self-Replication of Palindromic Duplex DNA." *Nature*, **369**(6477): 218-221.

56. Lee, D.H. *et al.* (1996). "A self-replicating peptide." *Nature*, **382**(6591): 525-528.

57. Hud, N.V. and Anet, F.A.L. (2000). "Intercalation-mediated synthesis and replication: a new approach to the origin of life." *J. Theor. Biol.*, **205**: 543-562.

58. Bean, H.D. *et al.* (2006). "Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA." *Orig. Life Evol. Biosph.*, **36**(1): 39-63.

59. Chaires, J.B. (1997). "Energetics of drug-DNA interactions." *Biopolymers*, **44**(3): 201-215.

60. Hurley, L.H. (1989). "DNA and associated targets for drug design." *J. Med. Chem.*, **32**(9): 2027-2033.

61. Hurley, L.H. (2002). "DNA and its associated processes as targets for cancer therapy." *Nature Reviews Cancer*, **2**: 188-200.

62. Krugh, T.R. (1994). "Drug-DNA interactions." *Curr. Opin. Struct. Biol.*, **4**(3): 351-364.

63. Lerman, L.S. (1963). "The Structure Of The DNA-Acridine Complex." *Proc. Nat. Acad. Sci. USA*, **49**: 94-102.

64. Siddiqui-Jain, A. *et al.* (2000). "Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription." *Proc. Nat. Acad. Sci. USA*, **99**(18): 11593-11598.

65. Kopka, M.L. *et al.* (1985). "The Molecular Origin of DNA Drug Specificity in Netropsin and Distamycin." *Proc. Nat. Acad. Sci. USA*, **82**(5): 1376-1380.

66. Berman, H.M. *et al.* (1979). "Molecular and Crystal-Structure of an Intercalation Complex - Proflavine-Cytidylyl-(3',5')-Guanosine." *Biopolymers*, **18**(10): 2405-2429.

67. Lerman, L.S. (1961). "Structural considerations in the interaction of DNA and Acridines." *J. Mol. Biol.*, **3**: 18-30.

68. Wang, A.H.J. *et al.* (1987). "Interactions Between An Anthracycline Antibiotic And DNA - Molecular Structure Of Daunomycin Complexed To d(CpGpTpApCpG) At 1.2-Å Resolution." *Biochemistry*, **26**(4): 1152-1163.

69. LePecq, J.B. and Paoletti, C. (1967). "A fluorescent complex between Ethidium Bromide and nucleic acids." *J. Mol. Biol.*, **27**: 87-106.

70. Nordmeier, E. (1992). "Absorption spectroscopy and dynamic and static light scattering studies of Ethidium Bromide binding to Calf Thymus DNA - Implications for outside binding and intercalation." *J. Phys. Chem.*, **96**(14): 6045-6055.

71. Qu, X.G. and Chaires, J.B. (2001). "Hydration changes for DNA intercalation reactions." *J. Am. Chem. Soc.*, **123**(1): 1-7.

72. Ren, J.S. and Chaires, J.B. (1999). "Sequence and structural selectivity of nucleic acid binding ligands." *Biochemistry*, **38**(49): 16067-16075.

73. Van dyke, M.W., Hertzberg, R.P. and Dervan, P.B. (1982). "Map of Distamycin, Netropsin, and Actinomycin binding sites on heterogeneous DNA - DNA cleavage inhibition patterns with Methidiumpropyl-EDTA.Fe(II)." *Proc. Nat. Acad. Sci. USA*, **79**(18): 5470-5474.

74. Gottesfeld, J.M. *et al.* (1997). "Regulation of gene expression by small molecules." *Nature*, **387**: 202-205.

75. White, S. *et al.* (1998). "Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands." *Nature*, **391**: 468-471.

76. Quintana, J.R., Lipanov, A.A. and Dickerson, R.E. (1991). "Low-Temperature Crystallographic Analyses of the Binding of Hoechst-33258 to the Double-Helical Dna Dodecamer C-G-C-G-a-a- T-T-C-G-C-G." *Biochemistry*, **30**(42): 10294-10306.

77. Rentzeperis, D. *et al.* (1995). "Interaction of Minor-Groove Ligands to an Aaatt/Aattt Site - Correlation of Thermodynamic Characterization and Solution Structure." *Biochemistry*, **34**(9): 2937-2945.

78. Wilson, W.D. *et al.* (1993). "The Search For Structure-Specific Nucleic Acid Interactive Drugs - Effects of Compound Structure On RNA vs. DNA Interaction Strength." *Biochemistry*, **32**(15): 4098-4104.

79. Hecht, S.M. (1986). "The chemistry of activated Bleomycin." *Acc. Chem. Res.*, **19**: 383-391.

80. Kane, S.A. *et al.* (1995). "Specific cleavage of a DNA triple helix by Fe-II center dot bleomycin." *Biochemistry*, **34**(51): 16715-16724.

81. Tan, J.D. *et al.* (1994). "NMR Evidence of Sequence-Specific DNA-Binding By a Cobalt(III)- Bleomycin Analog With Tethered Acridine." *Inorg. Chem.*, **33**(19): 4295-4308.

82. Sherman, S.E. *et al.* (1985). "X-Ray Structure Of The Major Adduct Of The Anticancer Drug Cisplatin With DNA - Cis-[Pt(NH3)2(d(pGpG))]." *Science*, **230**(4724): 412-417.

83. Arya, D.P. and Coffee, R.L. (2000). "DNA triple helix stabilization by aminoglycoside antibiotics." *Bioorg. Med. Chem. Lett.*, **10**(17): 1897-1899.

84. Helene, C. (1991). "The antigene strategy - control of gene expression by triplex forming oligonucleotides." *Anti Cancer Drug Design*, **6**(6): 569-584.

85. Jain, S.S. *et al.* (2004). "Enzymatic behavior by intercalating molecules in a template-directed ligation reaction." *Angew. Chem. Int. Ed.*, **43**(15): 2004-2008.

86. Jain, S.S., Polak, M. and Hud, N.V. (2003). "Controlling nucleic acid secondary structure by intercalation: effects of DNA strand length on coralyne-driven duplex disproportionation." *Nuc. Acid Res.*, **31**(15): 4608-4615.

87. Persil, O. *et al.* (2004). "Assembly of an antiparallel homo-adenine DNA duplex by small molecule binding." *J. Am. Chem. Soc.*, **126**(28): 8644-8645.

88. Polak, M. and Hud, N.V. (2002). "Complete disproportionation of duplex poly(dT).poly(dA) into triplex poly(dT).poly(dA).poly(dT) and poly(dA) by coralyne." *Nuc. Acid Res.*, **30**: 983-992.

89. Chan, L.M. (1970). "The interactions of aminoacridines with DNA." *Biochim. Biophys. Acta*, **204**: 252-254.

90. Deubel, V. and Leng, M. (1974). "Interaction between proflavine and double stranded polynucleotides." *Biochimie*, **56**: 641-648.

91. Lober, G. (1968). "On the complex formation of acridine dyes with DNA-IV. The equilibrium constants of substituted proflavine and acridine orange derivatives." *Photochem. Photobiol.*, **8**: 23-30.

92. Riemer, S.C. and Bloomfield, V.A. (1979). "Effect of Mg^{++} and polyamines on Proflavine binding to T2 DNA." *Biopolymers*, **18**(7): 1695-1708.

93. Drummond, D.S., Simpson-Gildemeister, V.F.W. and Peacocke, A.R. (1965). "Interaction of Aminoacridines with Deoxyribonucleic Acid: Effects of Ionic Strength, Denaturation, and Structure." *Biopolymers*, **3**(2): 135-153.

94. Ortona, O. *et al.* (1990). "Stacking Equilibria Of Proflavine In Various Solutions." *J. Mol. Liq.*, **45**(3-4): 201-211.

95. Quadrifoglio, F., Crescenzi, V. and Giancotti, V. (1974). "Calorimetry of DNA-Dye Interactions In Aqueous Solutions. 1. Proflavine And Ethidium Bromide." *Biophysical Chemistry 1.*: 319-324.

96. Felsenfeld, G. and Rich, A. (1957). "Studies on the formation of two and three-stranded polyribonucleotides." *Biochim. Biophys. Acta*, **26**: 457-468.

97. Stevens, C.L. and Felsenfeld, G. (1964). "The Conversion of Two Stranded Poly (A+U) to 3-Strand Poly (A+2U) and Poly A by Heat." *Biopolymers*, **2**(4): 293-314.

98. Hoogsteen, K. (1963). "Crystal Structure of 1-Methylthymine." *Acta Cryst.*, **16**(1): 28-38.

99. Hanvey, J.C., Shimizu, M. and Wells, R.D. (1988). "Intramolecular DNA triplexes in supercoiled plasmids." *Proc. Nat. Acad. Sci. USA*, **85**(17): 6292-6296.

100. Mirkin, S.M. *et al.* (1987). "DNA H-form requires a homopurine homopyrimidine mirror repeat." *Nature*, **330**(6147): 495-497.

101. Kamenetskii, M.D.F. and Mirkin, S.M. (1995). "Triplex DNA structures." *Annu. Rev. Biochem.*, **64**: 65-95.

102. Mergny, J.L. *et al.* (1991). "Intercalation of Ethidium-Bromide Into a Triple-Stranded Oligonucleotide." *Nuc. Acids Res.*, **19**(7): 1521-1526.

103. Mergny, J.L. *et al.* (1992). "Triple helix specific ligands." *Science*, **256**(5064): 1681-1684.

104. Scaria, P.V. and Shafer, R.H. (1991). "Binding of ethidium-bromide to a DNA triple helix - evidence for intercalation." *J. Biol. Chem.*, **266**(9): 5417-5423.

105. Duval-Valentin, G., Thuong, N.T. and Helene, C. (1992). "Specific-Inhibition of Transcription By Triple Helix-Forming Oligonucleotides." *Proc. Nat. Acad. Sci. USA*, **89**(2): 504-508.

106. Moser, H.E. and Dervan, P.B. (1987). "Sequence-Specific Cleavage of Double Helical DNA by Triple Helix Formation." *Science*, **238**(4827): 645-650.

107. Mooren, M.M.W. *et al.* (1990). "Polypurine Polypyrimidine Hairpins Form a Triple Helix Structure at Low Ph." *Nuc. Acids Res.*, **18**(22): 6523-6529.

108. Plum, G.E. (1997). "Thermodynamics of oligonucleotide triple helices." *Biopolymers*, **44**(3): 241-256.

109. Latimer, L.J.P. *et al.* (1995). "The binding of analogs of coralyne and related heterocyclics to DNA triplexes." *Biochem. Cell Biol.*, **73**(1-2): 11-18.

110. Lee, J.S., Latimer, L.J.P. and Hampel, K.J. (1993). "Coralyne binds tightly to both T·A·T-containing and C·G·C^+- containing DNA triplexes." *Biochemistry*, **32**(21): 5591-5597.

111. Wilson, W.D. *et al.* (1994). "The interaction of intercalators and groove-biding agents with DNA triple-helical structures: the influence of ligand structure, DNA backbone modifications and sequence." *J. Mol. Recog.*, **7**: 89-98.

112. Wilson, W.D. *et al.* (1993). "DNA triple-helix specific intercalators as antigene enhancers - unfused aromatic cations." *Biochemistry*, **32**(40): 10614-10621.

113. Kim, H.K. *et al.* (1996). "Interactions of intercalative and minor groove binding ligands with triplex poly(dA)center dot[poly(dT)](2) and with duplex poly(dA)center dot poly(dT) and poly[d(A-T)](2) studied by CD, LD, and normal absorption." *Biochemistry*, **35**(4): 1187-1194.

114. Wilson, W.D. *et al.* (1976). "Coralyne. Intercalation with DNA as a possible mechanism of antileukemic action." *J. Med. Chem.*, **19**: 1261-1263.

115. Moraru-Allen, A.A. *et al.* (1997). "Coralyne has a preference for intercalation between TA·T triples in intramolecular DNA triple helices." *Nuc. Acids Res.*, **25**(10): 1890-1896.

116. Xing, F.F. *et al.* (2005). "Molecular recognition of nucleic acids: Coralyne binds strongly to poly(A)." *FEBS Lett.*, **579**(22): 5035-5039.

117. *Circular Dichroism: Principles and Applications*. 2nd ed, ed. N. Berova, K. Naknishi, and R. Woody. 2000, New York: John Wiley & Sons. 877.

118. Arya, D.P. *et al.* (2001). "Aminoglycoside-nucleic acid interactions: remarkable stabilization of DNA and RNA triple helices by neomycin." *J. Am. Chem. Soc.*, **123**: 5385-5395.

119. Shi, X.C. and Chaires, J.B. (2006). "Sequence- and structural-selective nucleic acid binding revealed by the melting of mixtures." *Nuc. Acids Res.*, **34**(2).

120. Sandstrom, K. *et al.* (2002). "A-tract DNA disfavors triplex formation." *J. Mol. Biol.*, **315**: 737-748.

121. Fagan, P.A. *et al.* (1996). "An NMR study of [d(CGCGAATTCGCG)](2) containing an interstrand cross-link derived from a distamycin-pyrrole conjugate." *Nuc. Acids Res.*, **24**(8): 1566-1573.

122. Goodsell, D.S., Kopka, M.L. and Dickerson, R.E. (1995). "Refinement of Netropsin Bound to Dna - Bias and Feedback in Electron-Density Map Interpretation." *Biochemistry*, **34**(15): 4983-4993.

123. Garoff, R. *et al.* (2002). "Helical aggregation of cyanine dyes on DNA templates: effect of dye structure on homo- and heteroaggregates." *Langmuir*, **18**: 6330-6337.

124. Hopkins, H.P. *et al.* (1993). "Effects of C_2H_5OH, $Na^+_{(aq)}$, $N(CH_2CH_3)_4^+{}_{(aq)}$, and $Mg^{2+}_{(aq)}$ on the thermodynamics of double-helix-to-random-coil transitions of poly(dA)-poly(dT) and poly(dAdT)." *J. Chem. Thermodyn.*, **25**(1): 111-126.

125. Hopkins, H.P. *et al.* (1993). "Duplex and triple-helix formation with dA_{19} and dT_{19} - thermodynamic parameters from calorimetric, NMR, and circular-dichroism studies." *J. Phys. Chem.*, **97**(24): 6555-6563.

126. Guerrier-Takada, C. *et al.* (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." *Cell*, **35**: 849-857.

127. Gesteland, R. and Atkins, J.F., eds. *The RNA World, second edition: The nature of modern RNA suggests a prebiotic RNA world.* 1999, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 709.

128. Breaker, R.R. and Joyce, G.F. (1995). "Self-Incorporation Of Coenzymes By Ribozymes." *J. Mol. Evol.*, **40**(6): 551-558.

129. Baudoin, O. *et al.* (1998). "Stabilization of DNA triple helices by crescent-shaped dibenzophenanthrolines." *Chem. Eur. J.*, **4**(8): 1504-1508.

130. Xu, Y.Z. and Kool, E.T. (1997). "A novel 5'-iodonucleoside allows efficient nonenzymatic ligation of single-stranded and duplex DNAs." *Tet. Lett.*, **38**(32): 5595-5598.

131. Xu, Y.Z. and Kool, E.T. (1999). "High sequence fidelity in a non-enzymatic DNA autoligation reaction." *Nuc. Acid Res.*, **27**(3): 875-881.

132. von Hippel, P.H. and McGhee, J.D. (1972). "DNA-protein interactions." *Annu. Rev. Biochem.*, **41**: 231-300.

133. Bernal, J.D. *The Physical Basis of Life*. 1951, London: Routledge & Kegan Paul.

134. Cairns-Smith, A.G. *Genetic Takeover: And the Mineral Origins of Life*. 1982, Cambridge: Cambridge University Press.

135. Schwartz, A.W. (1996). "Did minerals perform prebiotic combinatorial chemistry?" *Chem. Biol.*, **3**(7): 515-518.

136. Kozlov, I.A. and Orgel, L.E. (2000). "Nonenzymatic template-directed synthesis of RNA from monomers." *Mol. Biol.*, **34**(6): 781-789.

137. Nelson, J.W., Martin, F.H. and Tinoco, I. (1981). "DNA And RNA oligomer thermodynamics - The effect of mismatched bases on double-helix stability." *Biopolymers*, **20**(12): 2509-2531.

138. Crick, F.H.C. (1966). "Codon-anticodon pairing: the wobble hypothesis." *J. Mol. Biol.*, **19**(2): 548-555.

139. Hunter, W.N. *et al.* (1986). "Refined crystal structure of an octanucleotide duplex with G.T mismatched base pairs." *J. Mol. Biol.*, **190**(4): 605-618.

140. Allawi, H.T. and SantaLucia, J. (1997). "Thermodynamics and NMR of internal G.T mismatches in DNA." *Biochemistry*, **36**(34): 10581-10594.

141. Ke, S.H. and Wartell, R.M. (1993). "Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA - Determination by temperature gradient gel electrophoresis." *Nuc. Acids Res.*, **21**(22): 5137-5143.

142. Peyret, N. *et al.* (1999). "Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches." *Biochemistry*, **38**(12): 3468-3477.

143. Helfgott, D.C. and Kallenbach, N.R. (1979). "Increased Binding of Ethidium-Bromide to Polynucleotide Duplexes Containing Mismatched Bases." *Nuc. Acids Res.*, **7**(4): 1011-1017.

144. Bhattacharya, P.K., Cha, J. and Barton, J.K. (2002). "H-1 NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches." *Nuc. Acids Res.*, **30**(21): 4740-4750.

145. Keniry, M.A., Owen, E.A. and Shafer, R.H. (1997). "The contribution of thymine-thymine interactions to the stability of folded dimeric quadruplexes." *Nuc. Acids Res.*, **25**(21): 4389-4392.

146. Koole, L.H., Vangenderen, M.H.P. and Buck, H.M. (1987). "A parallel right-handed duplex of the hexamer d(TpTpTpTpTpT) with phosphate triester linkages." *J. Am. Chem. Soc.*, **109**(13): 3916-3921.

147. Lee, S.J. and Hurley, L.H. (1999). "A thymine : thymine mismatch enhances the pluramycin alkylation site downstream of the TBP-TATA box complex." *J. Am. Chem. Soc.*, **121**(39): 8971-8977.

148. Lober, G. (1968). "On the complex formation of acridine dyes with DNA-IV. The equilibrium constants of substituted proflavine and acridine orange derivatives." *Photochem. Photobiol.*, **8**: 23-30.

149. Costantino, L. *et al.* (1984). "Acridine orange association equilibrium in aqueous solution." *J. Chem. Eng. Data*, **29**(1): 62-66.

150. Ortona, O. *et al.* (1990). "Stacking equilibria of proflavine in various solutions." *J. Mol. Liq.*, **45**(3-4): 201-211.

151. Van dyke, M.W., Hertzberg, R.P. and Dervan, P.B. (1982). "Map of distamycin, netropsin, and actinomycin binding sites on heterogeneous DNA - DNA cleavage inhibition patterns with methidiumpropyl-EDTA.Fe(II)." *Proc. Natl. Acad. Sci. USA*, **79**(18): 5470-5474.

152. Pelton, J.G. and Wemmer, D.E. (1990). "Binding modes of distamycin-A with d(CGCAAATTTGCG)2 determined By 2-dimensional NMR." *J. Am. Chem. Soc.*, **112**(4): 1393-1399.

153. Kapuscinski, J. and Szer, W. (1979). "Interactions of 4',6-diamidine-2-phenylindole with synthetic polynucleotides." *Nuc. Acids Res.*, **6**(11): 3519-3534.

154. Zuber, G. and Behr, J.P. (1994). "Nonenzymatic Plasmid Ligation-Mediated by Minor Groove-Binding Molecules." *Biochemistry*, **33**(26): 8122-8127.

155. Gilbert, M. and Claverie, P. (1968). "A theoretical study of electrostatic interactions in intercalation model of DNA-dye complex." *J. Theor. Biol.*, **18**(3): 330-349.

156. Drummond, D.S., Simpson-Gildemeister, V.F.W. and Peacocke, A.R. (1965). "Interaction of aminoacridines with deoxyribonucleic acid - Effects of ionic strength, denaturation, and structure." *Biopolymers*, **3**(2): 135-153.

157. Miller, G.P. and Kool, E.T. (2002). "A simple method for electrophilic functionalization of DNA." *Org. Lett.*, **4**(21): 3599-3601.

158. Freier, S.M. *et al.* (1983). "Effects Of 3' Dangling End Stacking On The Stability Of GGCC And CCGG Double Helices." *Biochemistry*, **22**(26): 6198-6206.

159. Marky, L.A., *Effect of phosphate groups on the stability of nucleic acid secondary structures*. 2006: personal communication.

160. Drummond, D.S. *et al.* (1966). "Interaction of Aminoacridines with Deoxyribonucleic Acid: Viscosity of the Complexes." *Biopolymers*, **4**: 971-987.

161. SantaLucia, J., Peyret, N. and Saro, P., *HyTher: Prediction of Nucleic Acid Hybridization Thermodynamics*. 1998, Chemistry Department, Wayne State University.

162. Sokolova, N.I. *et al.* (1988). "Chemical Reactions within DNA Duplexes - Cyanogen-Bromide As an Effective Oligodeoxyribonucleotide Coupling Agent." *FEBS Lett.*, **232**(1): 153-155.

163. Kim, S.K. *et al.* (1997). "Binding geometries of triple helix selective benzopyrido[4,3- b]indole ligands complexed with double- and triple-helical polynucleotides." *Biopolymers*, **42**(1): 101-111.

164. Marchand, C. *et al.* (2000). "A new family of sequence-specific DNA-cleaving agents directed by triple-helical structures: Benzopyridoindole-EDTA conjugates." *Chem. Eur. J.*, **6**(9): 1559-1563.

165. Pilch, D.S. *et al.* (1993). "Characterization of a Triple Helix Specific Ligand - BePI (3- Methoxy-7H-8-Methyl-1,1-[(3'-Amino)Propylamino]-Benzo[E]Pyrido[4,3-B]Indole) Intercalates Into Both Double Helical and Triple Helical DNA." *J. Mol. Biol.*, **232**(3): 926-946.

166. Yoshizawa, S. *et al.* (1997). "GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins." *Biochemistry*, **36**(16): 4761-4767.

167. Breaker, R.R. (1997). "DNA enzymes." *Nat. Biotech.*, **15**(5): 427-431.

168. Tuma, J. *et al.* (2004). "How much pi-stacking do DNA termini seek? Solution structure of a self-complementary DNA hexamer with trimethoxystilbenes capping the terminal base pairs." *Biochemistry*, **43**(50): 15680-15687.

169. Bjorndal, M.T. and Fygenson, D.K. (2002). "DNA melting in the presence of fluorescent intercalating oxazole yellow dyes measured with a gel-based assay." *Biopolymers*, **65**(1): 40-44.

VITA

Swapan Satyen Jain was born on August 13th, 1977 in Ludhiana, India, where he attended Atam Public School for his primary and senior secondary education. He began his BSc (Bachelor in Science) studies at Government College, Ludhiana in 1995. After coming to the USA in 1997, he transferred to Kennesaw State University, Kennesaw, Georgia where he finished with a Bachelor in Chemistry and a Bachelor in Biology in May 2001 (*summa cum laude)*. Swapan pursued his graduate studies in the School of Chemistry and Biochemistry at Georgia Institute of Technology in Atlanta, Georgia, USA.